Joint Image and Pupil Plane Reconstruction Algorithm based on
Bayesian Techniques

DISSERTATION

James D Phillips, Major, USAF

AFIT/DEE/ENG/08-07

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The views expressed in this dissertation are those of the author and do not reect the official policiy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/DEE/ENG/08-07

Joint Image and Pupil Plane Reconstruction Algorithm
based on Bayesian Techniques

DISSERTATION

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

James D Phillips, B.S.E.E., M.S.E.E.

Major, USAF

March 2008

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFIT/DEE/ENG/08-07

JOINT IMAGE AND PUPIL PLANE RECONSTRUCTION ALGORITHM
BASED ON BAYESIAN TECHNIQUES

James D Phillips, B.S.E.E., M.S.E.E.

Major, USAF

Approved:

_____   _____
Dr. Stephen C Cain, Ph.D.   25 Jan 2008
Dissertation Advisor   date

_____   _____
Dr. John O Miller, Ph.D.   25 Jan 08
Dean's Representative   date

_____   _____
Lt Col Matthew E Goda, Ph.D.   23 JAN 08
Committee Member   date

_____   _____
Maj Steven N Thorsen, Ph.D.   25 Jan 2008
Committee Member   date

_____   _____
Dr. Ed Watson, Ph.D.   18 JAN 2008
Committee Member   date

Accepted:

_____   _____
M. U. Thomas   7 Feb 2008
Dean, Graduate School of Engineering and Management   Date

AFIT/DEE/ENG/08-07

Abstract

The focus of this research was to develop a joint pupil and focal plane image recovery algorithm for use with coherent LADAR systems. The benefits of such a system would include increased resolution with little or no increase in system weight and volume as well as allowing for operation in the absence of natural light since the target of interest would be actively illuminated. Since a pupil plane collection aperture can be conformal, such a system would also potentially allow for the formation of large synthetic apertures.

The algorithm developed used many frames of coherent pupil and focal plane data. The data frames are summed in the respective planes to give two data sets (one for each plane). Appropriate statistical models are used and a joint Maximum Likelihood estimator is formed. The algorithm is tested using a Monte Carlo approach. The system is demonstrated to be robust and in all but extreme cases yields better results than algorithms using a single data set (such as deconvolution). It was shown that the joint algorithm had a resolution increase of 70% over deconvolution alone and a 40% increase over traditional pupil plane algorithms. It was also demonstrated that the new algorithm does not suffer as severely from stagnation problems typical with pupil plane algorithms. A stopping criteria based on the statistics of the data was also developed.

Acknowledgements

There are many people who deserve credit for assisting me along this educational path, but I would be remiss if I did not say "Thank You" first to my wife and children who supported me during this process. There were many nights that they waited patiently while I worked and could not give them the time that they deserve.

Next I would like to thank my committee, and especially my advisor, Dr Stephen Cain. Without the knowledge gained from each of you, this dissertation would not have been possible.

Above all I want to thank God to whom I owe everything.

James D Phillips

<div align="center">

Table of Contents

</div>

List of Figures

x

List of Symbols

Joint Image and Pupil Plane Reconstruction Algorithm based on Bayesian Techniques

I. Introduction

The focus of this research is to explore the problem of coherent image reconstruction using large synthetic aperture arrays. The benefits would include a large increase in system resolution for a small increase in system weight and volume. This type of imaging system would also allow for conformal optical detectors to be placed on existing platforms. Furthermore since the proposed system is an active imaging system, it would be able to be used when natural light was not available and would provide higher theoretical resolution than traditional IR systems.

1.1 Problem Definition

The two primary limiting factors for optical resolution are limited aperture size and phase error in the propagation paths. The phase error can be broken into two parts: 1) system static aberrations and 2) turbulence in the propagation path. There are many methods for characterizing and dealing with static system aberrations; we will propose a system design to mitigate the other two problems.

Since it is known that the aperture of an imaging system limits spatial resolution, it is desirable to form larger apertures. Large apertures have disadvantages as well; they are more difficult to manufacture, and due to weight, and size , they are not feasible for space-based applications [10]. By taking advantage of the coherence properties of laser light, it is possible to form a synthetic aperture array from many smaller, monolithic apertures. By doing this, one can expect to obtain higher spatial resolution than can be produced from existing monolithic apertures. Since it is difficult to recover absolute phase of an optical field, it is desirable to form the synthetic aperture without interfering light from the subapertures [11]. To accomplish this, a joint estimation algorithm using both pupil and image plane data will

1

be formed. The pupil plane data will be collected by the synthetic aperture while the image plane data will be collected from a reasonable sized monolithic aperture. The pupil plane estimation algorithm will be based on the correlography methods of Fienup and Idell [21], while the image plane algorithm will be derived from the deconvolution methods demonstrated by MacDonald [24]. These methods are combined in this research to form an improved image retrieval algorithm.

1.2 *Previous Image recovery work*

1.2.1 Deconvolution. Deconvolution is the process of estimating an unknown function, $f(x)$, from a noisy measurement of the convolution of $f(x)$ with a known function $h(x)$. If we have no knowledge of $h(x)$ it must also be estimated and the problem becomes blind deconvolution.

1.2.2 Phase Retrieval/Imaging Correlography. The term phase retrieval is used to describe any method of forming an image from the Fourier Modulus of the object field. The foundational work on phase retrieval was done by J.R. Fienup [7]. In this paper two iterative methods to recover an object from Fourier modulus data are shown; both methods are derived from the Gerchberg-Saxon algorithm [17]. The first of these methods is known as the error reduction approach and is shown in Figure 1.1a. In the error reduction approach the algorithm begins with a random guess of the object brightness function. This initial estimate is Fourier transformed and the modulus of the transform is replaced with the measured pupil plane data, which is then inverse Fourier transformed to form a new estimate. The new estimate is forced to comply with any known constraints in the object domain. These constraints include non-negativity of the object and any known object support. The term "object support" is used to refer to any area where the object is known to be non-zero. Since the Fourier domain data is related to the object autocorrelation, the diameter of the object support is related to the diameter of the Fourier domain data [5, 12]; the autocorrelation support will not give a unique solution for the object support, but
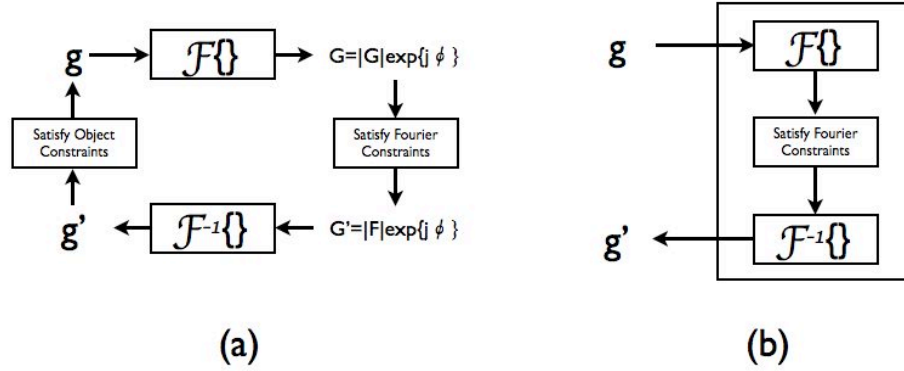
Figure 1.1: Fienup's phase retrieval algorithms (a) error reduction (b) input-output method [7]

the union of all possible object support sets can be found. This algorithm is run for a set number of iterations before exiting. In Figure 1.1b we see the input-output approach. This method will not be discussed, since the differences are minor and unimportant in this discussion. Other methods of phase retrieval, such as various gradient methods [8] and deconvolution techniques [33], can be found in the literature. Phase retrieval algorithms will often suffer from stagnation or uniqueness problems, however, efforts have been made to avoid these problems [4, 15].

Another method of forming images from pupil plane data is imaging correlography. This technique takes advantage of the fact that the autocorrelation of the object brightness function and the squared modulus of the Fourier transform of the object form a Fourier transform pair [2]. An iterative method of recovering images from correlations was shown by Schulz [32]. Again the details of Schulz's work are unimportant to this work other than to demonstrate that images can be recovered from autocorrelations with some degree of success.

Both of the above techniques benefit from an estimate of the object support. An algorithm that can directly measure a low resolution image and the pupil plane intensity was proposed by Fienup [14]. It is important to note he did not derive a

joint estimator from both data sets, rather he used the image as a support constraint, and the technique offered no method for dealing with atmospheric turbulence.

The data model this research uses is found in work by Idell [21]. The model show that the average of many realizations of the squared modulus of a speckled autocorrelation will converge ($\mathbb{R}$ convergence) to the convolution of the true autocorrelation and the PSF plus a dc term. Idell used this model and classic phase retrieval techniques to recover an image. This research differs in that it will use the same data model, but will form a maximum likelihood (ML) estimator.

1.2.3 Prior Synthetic Aperture efforts.   The method of image formation this research will propose is a lensless imaging technique; this means an image is not formed by the optics, but rather is calculated from the non-imaged pupil plane data. Imaging correlography is one method of doing this for incoherent light [13]. To form lensless coherent images other techniques must be employed. If the field in the pupil plane is measured using heterodyne detection this is a trivial problem of numerically propagating the field and forming a simulated image in a computer; however, heterodyne detection at optical wavelengths is challenging and has yet to be proven feasible for this application. For this reason other methods must be used. Fienup has proposed applying his phase retrieval algorithm along with a shaped illumination constraint to this problem [11]. This technique has been shown to work well with sharp-edged illumination patterns, but not as well for soft-edged illumination patterns [27]. The difficulty is that sharp-edged illumination patterns require large projection optics. If the large optics are available for beam projection they can also be used for imaging, neglecting the benefits of the lensless array.

This work is a continuation of initial studies done by Cain [3]. The work by Dr. Cain was a proof of concept that includes simplifications which will be removed. The differences in this work include, but are not limited to, using a more complete data model for the pupil plane data, and using a more accurate model for image plane statistics.

4

1.3   Significant Contributions

This work resulted in the following significant contributions:

1. Establish a maximum likelihood phase retrieval algorithm

2. Establish a joint estimation algorithm that is appropriate for use with coherent LADAR imagery

3. Demonstrate the algorithms are less sensitive to atmospheric turbulence that other methods of image recovery

4. Establish a new stopping/damping criteria for use with existing deconvolution algorithms which will avoid noise amplification.

1.4   Document Outline

The remainder of this document is organized as follows. Chapter 2 will discuss applicable optical theory while Chapter 3 will discuss estimation theory. In Chapter 4 a review of deconvolution is included, this chapter will also outline a statistics-based stopping criteria. Chapter 5 will demonstrate a new pupil plane algorithm. Finally chapter 6 will combine the algorithms of the previous two chapters and demonstrate a robust joint algorithm for image recovery.

II. Optical Theory

Active imaging techniques rely heavily on understanding fundamental properties of light including propagation, diffraction and coherence. This chapter is dedicated to providing an overview of these optical properties. Much of this material was adapted from [22, 23].

2.1 *Fourier Optics and Wave Propagation*

This section is devoted to giving the reader an understanding of Fourier optics and the physics of wave propagation. It will begin with a discussion of the propagation of monochromatic light and will conclude with a brief discussion of the properties of coherent and incoherent illumination. The propagation theory is drawn from [16, 20].

2.1.1 Wave Propagation - The monochromatic case. Armed with the knowledge that light can be modeled Electro-Magnetic wave, we can begin to formulate a model for propagation. We can write an expression for the electric field (for the remainder of the document the term field will be understood to mean electric field), $u(\xi, \eta, t)$, in the (ξ, η) plane as

$$u(\xi, \eta, t) = A(\xi, \eta) \cos[2\pi\nu t + \phi(\xi, \eta)] \tag{2.1}$$

where $A(\xi, \eta)$ is the field amplitude, ν is the optical frequency of the field, t is time, and $\phi(\xi, \eta)$ is the phase at position (ξ, η). For simplicity of notation we will define the complex amplitude, $U(\xi, \eta)$ of the field as

$$U(\xi, \eta) = A(\xi, \eta)e^{j\phi(\xi,\eta)} = |u(\xi, \eta)|e^{j\phi(\xi,\eta)} \tag{2.2}$$

For monochromatic light the propagation of the entire field can be accomplished by adding the appropriate phase delay to each point and summing the resultant phasors in the (x, y) plane. This is accomplished using the Rayleigh-Sommerfeld diffraction

integral

$$U(x,y) = \frac{z}{j\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\xi,\eta) \frac{e^{jkr}}{r^2} d\eta d\xi \tag{2.3}$$

where z is the normal distance between the input and output propagation planes, λ is the wavelength, $k = \frac{2\pi}{\lambda}$, and r is defined as

$$r = \sqrt{(x-\xi)^2 + (y-\eta)^2 + z^2} \tag{2.4}$$

2.1.2 Fresnel and Fraunhoffer Approximations. The Rayleigh-Sommerfeld formula, Equation 2.3, is computationally expensive (order N^4 for a N by N array) when implemented in digital simulations. For this reason, we will make some simplifications that can be applied provided that certain criteria are satisfied.

2.1.2.1 Fresnel Diffraction. The first of these simplifications is the Fresnel approximation [20]. Using the Maclaurin series expansion of the square root given by

$$\sqrt{1+b} = 1 + \frac{1}{2}b - \frac{1}{8}b^2 + ... \tag{2.5}$$

we can rewrite r from Equation 2.4

$$r \approx z[1 + \frac{1}{2}(\frac{x-\xi}{z})^2 + \frac{1}{2}(\frac{y-\eta}{z})^2] \tag{2.6}$$

eliminating terms of higher order than 1(in b). By taking the full form of Equation 2.6 in the exponential term and only the first term in the denominator term Equation 2.3 simplifies to

$$U(x,y) = \frac{e^{jkz}}{j\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(\xi,\eta) e^{\frac{jk}{2z}[(x-\xi)^2+(y-\eta)^2]} d\xi d\eta \tag{2.7}$$

By expanding the quadratic terms in the exponential this equation can be rewritten as

$$U(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{\lambda z}(x^2+y^2)} \mathcal{F}_2[U(\xi,\eta) e^{\frac{jk}{2z}(\xi^2+\eta^2)}] \tag{2.8}$$

where $\mathcal{F}_2$ is the 2-D Fourier transform defined by

$$\mathcal{F}_2[U(\xi,\eta)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(\xi,\eta)e^{-j2\pi(f_{(\xi)}\xi + f_{(\eta)}\eta)}d\xi d\eta \qquad (2.9)$$

evaluated at $f_\xi = \frac{x}{\lambda z}$ and $f_\eta = \frac{y}{\lambda z}$.

This approximation is valid provided the observation point is located far enough away from initial plane. This distance is given by [20]

$$z \gg \max_{(x,y)\in X_0,(\xi,\eta)\in \Psi_0} \sqrt[3]{\frac{\pi}{4\lambda}[(x-\xi)^2 + (y-\eta)^2]^2} \qquad (2.10)$$

where X_0 and Ψ_0 are variables in $\mathbb{R}^2$ that define the non-zero regions in the input and output planes respectively. The output plane will be limited by an aperture function to avoid having values everywhere.

2.1.2.2 Fraunhoffer Diffraction. The propagation integral shown in Equation 2.8 can be further simplified if one recognizes that the quadratic term inside the transfrom is approximately 1 for large z. This resulting equation is

$$U(x,y) = \frac{e^{jkz}}{j\lambda z}e^{\frac{jk}{\lambda z}(x^2+y^2)}\mathcal{F}_2[U(\xi,\eta)] \qquad (2.11)$$

evaluated at $f_x = \frac{x}{\lambda z}$ and $f_y = \frac{y}{\lambda z}$. Which means the output field is a simple Fourier transform of the input field. This is valid when

$$z \gg \max_{(\xi,\eta)\in \Psi_0} \frac{k(\xi^2 + \eta^2)}{2} \qquad (2.12)$$

It is desirable whenever possible to work in the Fraunhoffer region due to computational efficiency (order $N^2\log_2(N)$ for an NxN array when N is a power of 2 [28]) gained by using the Fourier transform and the fact that we can take advantage of many properties of the Fourier transform.
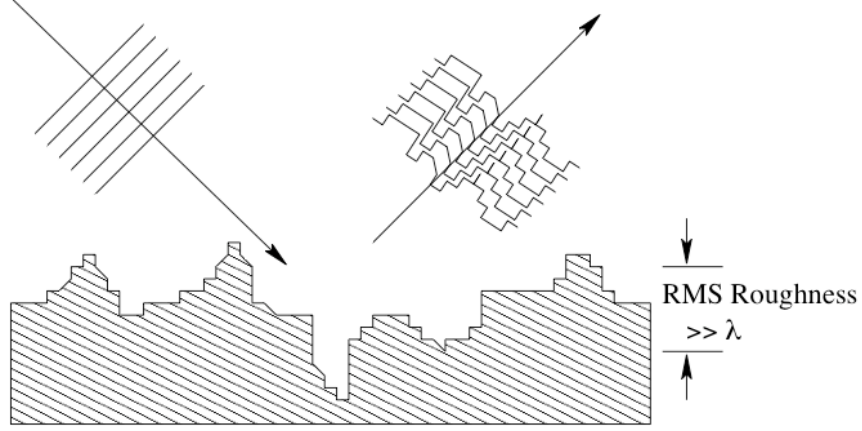
Figure 2.1:   Illustration of the phase aberration induced when coherent light is reflected from a optically rough surface

2.2   Speckle Statistics

The speckled appearance of coherent light scattered from a rough surface is a well known phenomenon. This speckle is caused by interference patterns in the detector plane. Since the surface roughness is random, so is the speckle pattern. When coherent light is reflected from a rough surface, the shape of the wavefront is changed in a random manner (see Figure 2.1). The wavefront deformation is modeled as a uniform random phase ($\phi \in (-\pi, \pi]$) added to the wavefront. Speckle has some useful statistical properties.

2.2.1   First Order Statistics.   This section is devoted to formulating a statistical model to describe the intensity and phase of propagated optical fields.

2.2.1.1   Assumptions.   All statistical models begin with assumptions; in our model phase and amplitude are considered random variables. First we assume the target is optically rough; "optically rough" surfaces have roughness (depth and breadth) on the order of the wavelength of the light. The roughness of these surfaces is modeled as a uniform random variable from $[-\pi, \pi]$. Second, we assume the target is illuminated by a plane wave; this assumption is not necessary but will simplify our math without a loss of generality. Third, we assume the phase and amplitude

9

of our field are independent of the other. Finally, we add the assumption that the phases and amplitudes of all contributing components are independent and identically distributed.

2.2.1.2 Development. The field at the receiver pupil is modeled as a linear combination off all the points in the target field:

$$A = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_i e^{j\theta_i} e^{j\phi_i} = ae^{j\theta} \tag{2.13}$$

where $\frac{a_i}{\sqrt{N}}$ is the magnitude of each field component, θ_i is the phase caused by the surface roughness, and ϕ_i is the phase associated with the atmospheric propagation. By writing the sum in this form, and including the scale factor of $\frac{1}{\sqrt{N}}$ we can use random phasor sums to proceed with the development. We are only concerned with the sum of these two phases, which when phase is represented modulo 2π, can be seen to possess a uniform distribution over $[-\pi, \pi]$. This is the same as the distribution on θ so we will drop ϕ for convenience. A is complex can also be written as

$$A = \alpha + j\beta \tag{2.14}$$

As N approached infinity, we can apply the central limit theorem and assume α and β are Gaussian. Therefore if we can calculate the mean and variance of α and β and their correlation coefficient, ρ, we can write the joint probability distribution function for α and β. We begin by finding the mean values

$$E[\alpha] = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} E[a_i] E[\cos \theta_i] = 0 \tag{2.15}$$

since the expectation is taken over $\theta_i \in [-\pi, \pi]$. By the same argument β is also zero mean. The variances are calculated as follows:

$$E[(\alpha - \bar{\alpha})^2] = E[\alpha^2] \tag{2.16}$$

$$E[\alpha^2] = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} E[a_i a_k \cos\theta_i \cos\theta_k] \tag{2.17}$$

Now for each i, choose $c_i \in \mathbb{R}$ such that $\theta_k = c_i\theta_i$ then

$$E[\alpha^2] = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} E[a_i a_k] E[\cos\theta_i \cos\theta_k] \tag{2.18}$$

$$E[\alpha^2] = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} E[a_i a_k] E[\cos\theta_i \cos c_i\theta_i] \tag{2.19}$$

using trigonometric identities can be written as

$$E[\alpha^2] = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} E[a_i a_k] E\left[\frac{\cos(\theta_i - c_i\theta_i) + \cos(\theta_i + c_i\theta_i)}{2} \right] \tag{2.20}$$

The second expectation is easily shown to be

$$E\left[\frac{\cos(\theta_i - c_i\theta_i) + \cos(\theta_i + c_i\theta_i)}{2} \right] = \begin{cases} \frac{1}{2} & c_i = 1 \rightarrow i = k \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

which yields a final result of

$$E[\alpha^2] = \frac{E[a_i^2]}{2} \tag{2.22}$$

In a similar fashion we can show that α and β are uncorrelated. The values calculated above allow us to write the joint pdf of α and β

$$p_{\alpha,\beta}(\alpha, \beta) = \frac{1}{\pi E[a_i^2]} \exp\left\{ -\frac{\alpha^2 + \beta^2}{E[a_i^2]} \right\} \tag{2.23}$$

We have come up with a joint pdf of the real and imaginary parts of the field in the pupil, but we want a marginal pdf on intensity, and if possible a marginal pdf for the phase. The next step to get there will be to transform $p_{\alpha,\beta}(\alpha, \beta)$ to a pdf on

11

amplitude and phase. The resulting joint pdf is

$$p_{A,\Theta}(a,\theta) = \frac{a}{\pi E[a_i^2]} e^{-\frac{a^2}{E[a_i^2]}} \tag{2.24}$$

by integrating out the other variable we can get marginal pdfs on phase and amplitude

$$p_A(a) = \frac{2a}{E[a_i^2]} e^{-\frac{a^2}{E[a_i^2]}} \tag{2.25}$$

$$p_\Theta(\theta) = \frac{1}{2\pi} \tag{2.26}$$

Finally we can find the pdf on intensity by noting that $I = A^2$ and doing another pdf transformation which yields

$$p_I(i) = \frac{1}{2\sigma^2} e^{-\frac{i}{2\sigma^2}} \tag{2.27}$$

This is a negative exponential distribution and has the property that $E[i] = 2\sigma^2$ which allow us to rewrite the pdf as

$$p_I(i) = \frac{1}{E[i]} e^{-\frac{i}{E[i]}} \tag{2.28}$$

The above argument applies to the pdf of instantaneous intensity, but since all detectors have a finite integration time we are interested in the statistics of integrated intensity. The reader is referred to chapter 6 of reference [19] for a more complete development. The resulting pdf of integrated intensity is

$$p_W(W) = \left(\frac{\mathcal{M}}{\bar{W}}\right)^{\mathcal{M}} \frac{W^{\mathcal{M}-1} exp\left(-\mathcal{M}\frac{W}{\bar{W}}\right)}{\Gamma(\mathcal{M})} \tag{2.29}$$

where W is the integrated intensity, $\bar{W}$ is the expected value of the mean intensity, and $\mathcal{M}$ is the speckle parameter of the light. By recognizing that there will be detection noise we can transform this PDF one final time. The detection noise is poisson and

12

the final pdf for the number of photocounts in the integration time is given below:

$$p(K) = \frac{\Gamma(K + \mathcal{M})}{\Gamma(K+1)\Gamma(\mathcal{M})} \left[1 + \frac{\mathcal{M}}{\overline{K}}\right]^{-K} \left[1 + \frac{\overline{K}}{\mathcal{M}}\right]^{-\mathcal{M}} \qquad (2.30)$$

where Γ is the well known Γ-function. This is the negative binomial distribution and accurately models photocount statistics [19].

2.3   Coherent Imaging

This section will develop a model of coherent imaging and contrast it with incoherent imaging. An understanding of linear systems will be needed and will therefore be provided up front.

2.3.1   Linear Systems Review.   In order to understand linear system analysis we must first define what we mean by a system. A system is defined as a mapping of a set of input functions to a set of output functions [20]. For optical imaging problems the set of input and output functions can represent either real-valued intensity or complex-valued field amplitude; in either case the functions are defined in 2-D variable space.

A convenient way to represent a system is as a mathematical operator, $\mathcal{S}\{\}$, which will operate on a 2-D function to produce an output that is also a 2-D function

$$g_2(x, y) = \mathcal{S}\{g_1(\xi, \eta)\} \qquad (2.31)$$

It is important to note that this relation can be many to one or one to one, but since for now we will limit our scope to deterministic systems it can not be a one to many mapping.

Now that we have defined a system, we must proceed to define the more restrictive case of *linear* systems. A system is referred to as linear if the superposition

13

principle is obeyed for all input functions p and q and all complex constants a and b

$$\mathcal{S}\{ap + bq\} = a\mathcal{S}\{p\} + b\mathcal{S}\{q\} \tag{2.32}$$

The advantage of linearity is the ability to express the output of a system in terms of a sum of "decomposed" inputs. To further this idea we can look at the response of a system to a displaced delta function, $\delta(x_1 - \xi, y_1 - \eta)$

$$h(x, y; \xi, \eta) = \mathcal{S}\{\delta(x - \xi, y - \eta)\} \tag{2.33}$$

The function h is referred to as the impulse response of the system. This allows us to relate the input and output by

$$g_2(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(\xi, \eta) h(x, y; \xi, \eta) d\xi d\eta \tag{2.34}$$

From here we will further restrict our interest to linear *shift-invariant* systems. A shift-invariant system is a system whose impulse response is only dependent on the separation of points in space and not the points themselves. For such systems we can write the impulse response as

$$h(x, y; \xi, \eta) = h(x - \xi, y - \eta) \tag{2.35}$$

Using this result we rewrite the Equation 2.36 as

$$g_2(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta \tag{2.36}$$

which is recognized to be a 2-D convolution of the input function with the impulse response. For convenience in later chapters we define short hand notation for convolution as

$$g_2(x, y) = [g_1 * h](x, y) \tag{2.37}$$

14

2.3.2   Imaging - A linear systems approach.   Equation 2.7 can be rewritten as

$$U_i(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x-\xi, y-\eta) U_o(\xi, \eta) d\xi d\eta \tag{2.38}$$

where U_i is the field in the image plane, U_o can be thought of as a geometric prediction of the field in the image plane and $h(u,v) = \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}[u^2+v^2]}$. This is easily recognized to be a convolution. By writing the propagation process as a convolution we can infer that the process is linear in complex field and space invariant [16]. For a coherent system the instantaneous intensity in the receiver plane can be found by

$$I_i(x,y) = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x-\xi, y-\eta) U_o(\xi, \eta) d\xi d\eta \right|^2 \tag{2.39}$$

As mentioned before, all detectors integrate for a time period and therefore average resulting in

$$E[I_i(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E[U_o(\xi_1, \eta_1) U_o^*(\xi_2, \eta_2)] h(x-\xi_1, y-\eta_1) h^*(x-\xi_2, y-\eta_2) d\xi_1 d\eta_1 d\xi_2 d\eta_2 \tag{2.40}$$

where the expectation is taken over the phase of U_o. However we know from above that for a single speckle realization this is a deterministic value and the expectation can be dropped.

2.3.2.1   Incoherent Imaging Systems.   It is necessary to take a brief look at incoherent imaging systems. The expectation in Equation 2.40 represents the mutual intensity. We have used the time average and ensemble averages interchangeably here since the random parameter is a time varying phase value. For incoherent light this mutual intensity is

$$E[U_o(\xi_1, \eta_1) U_o^*(\xi_2, \eta_2)] = \delta(\xi_1 - \xi_2, \eta_1 - \eta_2) I_o(\xi_1, \eta_1) \tag{2.41}$$

Using this result along with Equation 2.40 we are able to write an expression relating the the object and image intensities

$$I_i(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x - \xi, y - \eta) I_o(\xi, \eta) \tag{2.42}$$

where the point spread function, s, is given by

$$s(x, y) = |h(x, y)|^2 \tag{2.43}$$

It is important to recall that I_o is not the true intensity distribution of the object, but rather a geometric image of the object in the image plane.

2.3.2.2 Multiframe Coherent Imaging. Incoherent imaging is important due to the fact that as many independent speckle realizations are imaged and summed in a coherent system the result approaches that predicted by an incoherent system [21].

2.4 Imaging through turbulence

The above propagation theory applies to light propagating in a medium with uniform index of refraction. In the atmosphere this is not the case. The consequence of this is that atmospheric turbulence becomes the limiting factor for resolution for most optical systems that require a long propagation through the atmosphere. In long exposure imaging, the point spread function (PSF) of the imaging system is very broad and smooth; while in short exposure imaging, known as speckle imaging, the PSF is not quite as broad but suffers from a modulated (speckled) irradiance pattern. In either of these two cases angular resolution is severely limited [31].

Turbulence effects result from random spatial and temporal fluctuations in index of refraction in the atmosphere, which in turn cause a random variation in optical path length (OPL). These variations in OPL result in phase abberations on the wavefront, which in turn become intensity variations after the wave has propagated. Since

atmospheric turbulence is not easily modeled as a deterministic process, statistical models are required to understand and model these effects. The first of these models was created in the 1940's by A.N. Kolmogorov.

The atmosphere can be considered a viscous fluid, and therefore it has two distinct states of motion - laminar and turbulent. The distinction between these two states is that laminar flow is smooth and regular while turbulent flow is unstable and acquires random subflows called turbulent eddies. The separation between these two regimes is defined by the Reynolds number:

$$Re = \frac{v_{avg}l}{k_v} \tag{2.44}$$

where v_{avg} is the average air velocity, l is the scale size, and k_v is the viscosity of the air. When the Reynolds number exceeds some critical value the flow is said to be turbulent. As an example, the viscosity of air is $k_v = 1.5 \times 10^{-5} \frac{m^2}{s}$, and assuming a scale size of $l = 10m$ and a velocity of $v_{avg} = 1\frac{m}{s}$, a Reynolds number of 6.7×10^5 is found. This example demonstrates that atmospheric air flow is essentially always turbulent [31].

In Kolmogorov's theory, he suggested the structure of the atmosphere, for large Reynolds numbers, was homogenous and isotropic within the inertial subrange. Inside the inertial subrange the atmosphere is comprised of eddies that interact and exchange energy to form and divide into smaller eddies. An eddy is defined as a pocket of air that has a uniform temperature and pressure [31]. The inertial subrange is defined by eddy sizes bounded by the inner scale, l_0, and the outer scale, L_0.

Index of refraction variations in the atmosphere result from temperature inhomogeneities caused by turbulent air motion. Since temperature fluctuations are a function of location in space, $\mathbf{R}$, and time, t, so is the index of refraction:

$$n(\mathbf{R}, t) = n_0 + n_1(\mathbf{R}, t), \tag{2.45}$$

where $n_0 \approx 1$ is the mean value of the index of refraction and $n_1(\mathbf{R}, t)$ is the deviation about this mean. The index of refraction time dependence can be ignored since the rate of change of the atmosphere is slow when compared to the typical timescales of turbulence moving across the beam (Taylors Frozen Flow) [1]. Using these simplifications, the index of refraction can be represented by

$$n(\mathbf{R}) = 1 + n_1(\mathbf{R}), \tag{2.46}$$

Using Equation 2.46, it is possible to arrive at the structure function describing the index of refraction variations in the atmosphere.

Because it is not possible to exactly describe the index of refraction random process for all positions in space, the structure function is necessary. There are too many random behaviors and variables to account for in a closed form solution. The index can only be described in reference to stationary random functions. Over long spatial periods, the index of refraction is not a stationary random process, but over short spatial periods of interest to applications of laser propagation, the index is considered to have stationary increments [1]. In other words, it is possible to treat the index random process as stationary with emphasis on the function $n(\mathbf{R}+\mathbf{R}_1)-n(\mathbf{R}_1)$. Intuitively, the structure function of the index of refraction is the mean squared difference between the index of refraction at one point in space and the index at a point with some separation distance from the first point. The structure function of $n(\mathbf{R})$ is defined by:

$$D_n(\mathbf{R_1}, \mathbf{R_2}) = \langle [n(\mathbf{R_1}) - n(\mathbf{R_2})]^2 \rangle \tag{2.47}$$

where $\mathbf{R_1}$ and $\mathbf{R_2}$ are vectors describing points in space and $\langle \cdot \rangle$ denotes the ensemble average. By starting with the structure function of wind velocity Kolmogorov was able to determine the structure function of the index of refraction to be:

$$D_n(R) = \begin{cases} C_n^2 R^{2/3} & , \quad l_o \ll R \ll L_o \\ C_n^2 l_o^{-4/3} R^2 & , \quad\quad R \ll l_o \end{cases} \tag{2.48}$$

18

where l_0 and L_0 are the inner and outer scale sizes, $R = |\mathbf{R_2} - \mathbf{R_1}|$, and C_n^2 is the atmospheric structure constant. At small scale sizes below l_o, the structure function follows a squared relationship (second part of Equation 2.48) which is found by performing a Taylor Series expansion on the structure function for small separation distances [1]. The structure function is dependent on the separation distance R and has units of radians squared; it can be written in terms of the atmospheric Fried parameter:

$$D_n(R) = 6.88 \left(\frac{R}{r_o} \right)^{5/3}, \tag{2.49}$$

where r_o relates to turbulence strength and is defined and discussed in a later section.

The structure function is related to the autocorrelation function of the index of refraction, Γ_n, by:

$$D_n(R) = 2[\Gamma_n(0) - \Gamma_n(R)] \tag{2.50}$$

Further the autocorrelation function, when it exist, is related to the power spectral density (PSD) , $\Phi_n(\kappa)$ by the *Wiener-Khinchin theorem* :

$$\Gamma_n(\mathbf{R}) = \int_{-\infty}^{\infty} \Phi_n(\vec{\kappa}) e^{j\vec{\kappa} \cdot \mathbf{R}} d\vec{\kappa} \tag{2.51}$$

From this relation a spectral model for the atmosphere can be developed.

The statistical distribution of size and number of turbulent eddies is described by the PSD of n, $\Phi_n(\vec{\kappa})$ where $\vec{\kappa}$ is the spatial wavenumber vector. The PSD can be thought of as a measure of the relative abundances of turbulent eddies at a given scale size. Using the assumption that the index of refraction is homogenous and isotropic, the PSD can be written as a function of the scalar wavenumber, κ [31]. Kolmogorov's theory only predicts a form for the PSD inside the inertial subrange:

$$\Phi_n(\kappa, z) = 0.033 C_n^2(z) \kappa^{\frac{-11}{3}} \tag{2.52}$$

where $C_n^2(z)$ is the structure constant of the atmosphere as a function of location in the propagation path, z.

The Kolmogorov spectrum is not valid for all wavenumbers so a more complete model is required. For a more complete model the *modified atmospheric spectrum* will be used and is given by [1]:

$$\Phi_n(\kappa, z) = 0.033 C_n^2(z) \left[1 + 1.802\,(\kappa/\kappa_l) - 0.254\,(\kappa/\kappa_l)^{7/6}\right] \frac{\exp\left(-\kappa^2/\kappa_l^2\right)}{(\kappa^2 + \kappa_o^2)^{11/6}} \quad, \quad 0 \leq \kappa \leq \infty$$
(2.53)

where $\kappa_l = \frac{3.3}{l_o}$ and $\kappa_0 = \frac{1}{L_o}$. The atmospheric model can be tailored by selecting appropriate inner scale, outer scale, and C_n^2 values depending on the laser beam propagation scenario.

The strength of turbulence in the atmosphere, C_n^2, depends on height above ground and the model chosen. Total turbulence strength for the entire path is found by integrating $C_n^2(z)$ over the path that laser light would travel to the sensor. To accomplish this task, the Hufnagel-Valley (H-V) model is chosen describing C_n^2. Like the modified spectrum for the atmosphere, the H-V model is most commonly used for generic conditions describing C_n^2, as it is based on real data of various seasons, altitudes, and geographic locations [1]. The H-V model used is

$$C_n^2(\mathrm{h}) = 0.00594(v/27)^2(10^{-5}\mathrm{h})^{10} \exp(-\mathrm{h}/1000) + 2.7 \times 10^{-16} \exp(-\mathrm{h}/1500)$$
$$+ A\exp(-\mathrm{h}/100)$$
(2.54)

where h is the height above the ground, v is the root-mean-square wind speed in (m/s) and A is the value of $C_n^2(0)$ at the ground in $m^{-2/3}$.

2.4.1 Atmospheric Parameters. Three atmospheric parameters are often used to describe turbulence strength: the Fried parameter, r_0; the isoplanatic angle, θ_0; and the Rytov variance, σ_1^2. Each of these parameters is a different moment of C_n^2

and is defined below:

$$r_0 = 1.67 \left[k^2 \int_0^L C_n^2(z) dz \right]^{-3/5} \tag{2.55}$$

$$\theta_0 = \left[2.91 k^2 \int_0^L z^{5/3} C_n^2(z) dz \right]^{-3/5} \tag{2.56}$$

$$\sigma_1 = 2.25 k^{7/6} \int_0^L C_n^2(z) z^{5/6} dz \tag{2.57}$$

where $z = \frac{h}{\cos \theta_z}$ and θ_z is the zenith angle. The Fried parameter defines the roll off of the OTF of the atmosphere [31]; another way of saying this is that little is gained in resolution for aperture sizes larger than r_0. In this research r_0 is the parameter we will use to describe turbulence strength.

2.5 Phase Screens

For long propagations through a non-uniform media it is necessary to have a way of modeling the phase perturbations as discrete layers of phase that can be added to the unperturbed wave. This type of model is called a phase screen. Depending on the effects to be modeled and the level of accuracy required one phase screen may or may not be sufficient. If more than one phase screen is used the strength of each screen must be adjusted accordingly. If r_0 is used to define the strength of the atmosphere then each phase screen can be assigned a strength according to

$$r_0^{-5/3} = \sum_{i=1}^N r_{0i}^{-5/3} \tag{2.58}$$

where r_{0i} is the Fried parameter of the individual phase screens [31, 72].

2.6 Phase Screen Creation

Modeling the atmosphere using knowledge of the scenario and power spectrum allows phase screens to be produced to represent the atmospheric random process. Several methods exist for producing phase screens using the power spectrum. Two common methods involve using the Zernike polynomial basis set to produce phase

screens and using the inverse Fourier transform of the power spectrum with Gaussian random variables to produce phase screens. One disadvantage in the Fourier transform method is that for simulations over longer time periods, an increasingly large screen must be computed. Additionally, the Fourier transform method produces screens that lack low frequency accuracy. In other words, the modified spectrum to be modeled contains a large percentage of power in the low frequency components. In taking the inverse Fourier transform, these low frequency regions are not allocated enough samples, so low frequencies are under-represented. To alleviate these two problems, a modification to the Fourier transform method is used called the "generalized Fourier series method " [26].

To facilitate understanding the Fourier series method, the power spectrum must be discussed in relation to random processes. Intuitively, the power spectrum of a random process is the average amount of power in each frequency component composing the random process. For this case, the random process is phase variation induced by the atmosphere. The power spectrum is related to the covariance of the phase variation, $B_n(R)$ by the *Wiener-Khintchine theorem* (recall Equation 2.51).

Starting with the modified power spectrum representing the atmosphere (Equation 2.53), one can finely sample the PSD the low frequency region and then give fewer samples to the high frequency region. In this way, the frequency regions that have a larger power are sampled more often. The PSD is then randomized using Gaussian variables with the appropriate variance. The result is an array of complex coefficients describing the frequency composition of a phase screen iteration. The complex coefficients exhibit circular complex Gaussian statistics with a variance corresponding to the previously sampled PSD. All that remains is to sum sinusoids of corresponding frequencies to produce the desired phase screen. Implementing this procedure is accomplished by the inverse Fourier series given by

$$\phi_k(x,y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbf{c}_{n,m} e^{j2\pi(f_{x_n}x + f_{y_m}y)}, \tag{2.59}$$

where (x, y) are spatial coordinates of the screen, $\mathbf{c}_{n,m}$ are randomized complex coefficients from the PSD of interest, and (f_{x_n}, f_{y_m}) are spatial frequency components from the PSD. Note that the sum is calculated rather than using a Fast Fourier Transform (FFT); this is necessary due to the nonlinear sampling method. By randomizing the real part of the PSD using a circular complex Gaussian random variable, complex coefficients $\mathbf{c}_{n,m}$ are created containing a random phase. Therefore, each phase screen iteration $\phi_k(x, y)$ is unique and possesses a unique random phase in the Fourier domain.

An advantage in this method appears for applications requiring a sequence of screens to represent longer time periods (several seconds). Instead of calculating one large phase screen and moving the area of interest around the screen as time progresses, it is only necessary to calculate the screen exactly where it is needed. Although the generalized Fourier series method cannot take advantage of fast Fourier transform algorithms, calculations are still saved by only calculating the screen area of interest. To implement the Fourier series method, a PSD for the turbulence of interest is calculated using l_0, L_0, r_0 and the spatial frequency region of concern. The Fourier series coefficients are then calculated for frequency components of interest (calculating more low frequency components). Afterward, the coefficients are used to construct the phase screen by summing sinusoids with different weights at any location desired.

III. Estimation Thoery

Estimation theory is a well defined method for inferring values of unknown parameters based on available observations which are corrupted by noise or incomplete. The estimate is formulated based on a cost function that weights the penalty for incorrect guesses. The basic estimation problem has four main components: a parameter space, a probabilistic mapping, an observation space and an estimation rule (see Figure 3.1).

The parameter space is made up of all possible parameters values, including the set that corresponds to reality. The parameter values generated by the estimation rule are also contained in the parameter space and hopefully lie "close" to the "true" data. The probabilistic mapping is a statistical model of the process (or processes) that describes how uncertainty is incorporated in the model for the data. The mapping is posed in the form of probability distribution functions of the data given the observations. The observation space is composed of all possible observations and may or may not intersect the parameter space. Finally, the estimation rule is the mapping used to map elements of the observation space to elements of the parameter space.

The parameter and observation spaces as well as the probabilistic mapping are all determined by the problem to be solved and cannot be changed. The task for the designer is to develop an estimation rule that gives acceptable results. The theory on estimation in this chapter is taken from [29, 34]

3.1   Bayes Estimation

If the parameter to be estimated, a, is a random variable one could choose Bayesian estimation as a means to develop an estimation rule. The first step in this estimation procedure it to define a cost function. The cost function will be used to weight estimates, $\hat{a}$. The cost function, $C(a, \hat{a}(\mathbf{R}))$ ($\mathbf{R}$ is a vector of observations), is a function of the true data and the estimate; however, it is usually sufficient to write
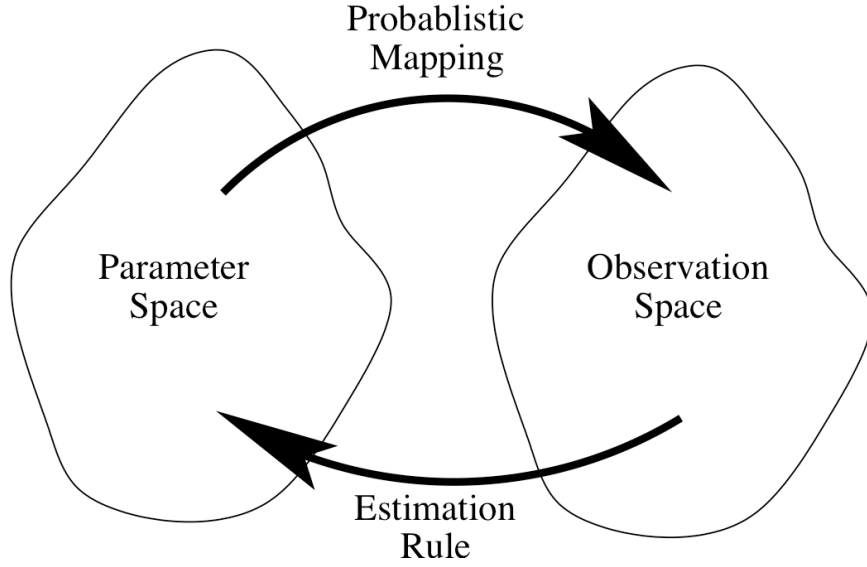
Figure 3.1: Diagram showing relationships between parameter and observation spaces and mapping/decision rules

the cost as a function only of the error of the estimate, a_ϵ

$$a_\epsilon = \hat{a}(\mathbf{R}) - a \tag{3.1}$$

This cost function, $C(a_\epsilon)$, is a function of a single variable and is more convenient to work with.

Three common cost functions are the squared error, the absolute error, and the uniform cost function. The squared error cost function,

$$C(a_\epsilon) = a_\epsilon^2, \tag{3.2}$$

clearly emphasizes large errors. The absolute error cost function,

$$C(a_\epsilon) = |a_\epsilon|, \tag{3.3}$$

is directly proportional to the magnitude of the error. Finally the uniform cost function,

$$C(a_\epsilon) = \begin{cases} 0 & |a_\epsilon| < \frac{\Delta}{2} \\ 1 & |a_\epsilon| \geq \frac{\Delta}{2} \end{cases}, \tag{3.4}$$

gives no weight to the magnitude of the error, but rather gives a uniform penalty if the error surpasses a threshold, Δ. The cost function is chosen by the designer based on the problem; for example a tracking system might chose the squared error cost function to penalize extreme errors, while a targeting system would choose a uniform cost function since anything outside of a certain error is unacceptable.

When using Bayesian estimation, it is necessary to have a known or estimated *a priori* probability distribution (prior) for the variable to be estimated; later section will discuss how to deal with estimation problems when the prior is unknown. Once we have selected a cost function, an expression for the risk is formed

$$\mathcal{R} \equiv E\{C[a, \hat{a}((\mathbf{R})]\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C[a, \hat{a}((\mathbf{R})]p_{A,\mathbf{R}}(a, \mathbf{r}) \tag{3.5}$$

The Bayes estimate is simply the estimate that minimizes the risk function. The difficulty is that the joint probability, $p_{A,\mathbf{R}}(a, \mathbf{r})$, is normally not available. To overcome this we recognize that the joint probability can be rewritten as

$$p_{A,\mathbf{R}}(a, \mathbf{r}) = p_{\mathbf{R}}(\mathbf{r})p_{A|\mathbf{R}}(a|\mathbf{r}) \tag{3.6}$$

This is further complicated by the fact that $p_{A|\mathbf{R}}(a|\mathbf{r})$ is also normally not defined, however it can be found via Bayes rule

$$p_{A|\mathbf{R}}(a|\mathbf{r}) = \frac{p_{\mathbf{R}|A}(\mathbf{r}|a)p_A(a)}{p_{\mathbf{R}}(\mathbf{r})} \tag{3.7}$$

We will see later it is not necessary to have $p_{\mathbf{R}}$.

This leads to the important question of how to handle problems where the prior of the parameter to be estimated, $p_A(a)$, is unknown. The next section on maximum likelihood estimation addresses how to approach this problems.

3.2   *Maximum Likelihood Estimation*

Recall from above that $\mathbf{R}$ is a vector of observations and A is the true value we are attempting to estimate. If we have no knowledge of the distribution function of A, the method in the previous section cannot be used. Instead we choose as our estimate the value of A that most likely led to the observed values of $\mathbf{R}$. The first step in this process is to define the log-likelihood function

$$\mathcal{L}(a) = \ln \left[p_{\mathbf{R}|A}(\mathbf{r}|a) \right] \tag{3.8}$$

The maximum likelihood estimate is the value of a where this function is maximized. If the maximum exist in the range of a, and $\mathcal{L}(a)$ has a continuous first derivative, then $\hat{a}_{ml}$ is found by

$$\left. \frac{d\mathcal{L}(a)}{da} \right|_{a=\hat{a}_{ml}} = 0 \tag{3.9}$$

Often this derivative is difficult to solve in closed form and we must resort to iterative methods; one such method is the Richardson-Lucy Algorithm and it is discussed in the next section.

3.3   *Richardson-Lucy Algorithm*

The Richardson-Lucy [30] algorithm is a modified gradient ascent method used for finding the maximum of Poisson likelihood functions, where $\frac{d\mathcal{L}(a)}{da}$ is the gradient of the likelihood function $\mathcal{L}(a)$. A typical gradient ascent algorithm begins by assuming $\mathcal{L}(a)$ is maximum at a_{max}. Next an arbitrary starting point, a_0, is chosen and the value of $\frac{d\mathcal{L}(a)}{da} \mid_{a=a_0}$ is found; if this value is positive then $a_{max} > a_0$, if it is negative

then $a_{max} < a_0$. We now calculate a_1 according to

$$a_n = a_{n-1} + \kappa \frac{d\mathcal{L}(a)}{da} \Big|_{a=a_n-1} \tag{3.10}$$

This algorithm iterates until $\frac{d\mathcal{L}(a)}{da} \big|_{a=a_n} \approx 0$. The positive constant, κ, must be sufficiently small to allow for convergence.

The Richardson-Lucy algorithm performs gradient ascent by breaking $\frac{d\mathcal{L}(a)}{da}$ into positive and negative parts

$$\frac{d\mathcal{L}(a)}{da} \Big|_{a=a_n} = \frac{d\mathcal{L}(a)^+}{da} \Big|_{a=a_n} + \frac{d\mathcal{L}(a)^-}{da} \Big|_{a=a_n} \tag{3.11}$$

It should be clear that if $\frac{d\mathcal{L}(a)^+}{da} \big|_{a=a_n} > \frac{d\mathcal{L}(a)^-}{da} \big|_{a=a_n}$ then $\frac{d\mathcal{L}(a)}{da} \big|_{a=a_n}$ is positive and therefore $a_{max} > a_n$, conversely if $\frac{d\mathcal{L}(a)^+}{da} \big|_{a=a_n} < \frac{d\mathcal{L}(a)^-}{da} \big|_{a=a_n}$ then $\frac{d\mathcal{L}(a)}{da} \big|_{a=a_n}$ is negative and $a_{max} < a_n$. The R-L algorithm forms a ratio of $\frac{d\mathcal{L}(a)^+}{da} \big|_{a=a_n} : \frac{d\mathcal{L}(a)^-}{da} \big|_{a=a_n}$ and updates a_n according to

$$a_n = a_{n-1} \frac{\frac{d\mathcal{L}(a)^+}{da} \big|_{a=a_{n-1}}}{\frac{d\mathcal{L}(a)^-}{da} \big|_{a=a_{n-1}}} \tag{3.12}$$

3.4 Estimator Quality

Once we have established an estimation routine, we would like to determine the quality of our estimator. To do this we would attempt to find the bias and the variance of the estimator. The bias of our estimator is defined by the following equation

$$B(a) = E\left[\hat{a}(\mathbf{R}) - a\right] \tag{3.13}$$

If the bias is zero we say we have an unbiased estimator, if it is non-zero and not a function of a we have a known bias, if it is a function of a we have an unknown bias.

For an iterative Richardson-Lucy (R-L) algorithm, the bias cannot be directly calculated. To find the bias for this estimator, we take advantage of the fact that for Poisson statistics the R-L algorithm converges to the maximum likelihood solution

28

[35]. If we can show that a maximum likelihood estimate is unbiased, the R-L estimate will also be unbiased.

Even an unbiased estimator can give unacceptable results for a particular trial; the estimator may indeed have a pdf centered on the true value, but a large variance. It is desirable for an estimator to have a small variance in the estimated values. It is often difficult to calculate the variance of an estimator; it is usually much easier to calculate a lower bound on the estimator variance and then compare the actual performance of the algorithm to this lower bound. One lower bound for unbiased estimators is the Cramér-Rao bound defined by

$$\text{Var}\left[\hat{a}(\mathbf{R}) - a\right] \geq \left(E\left\{ \left[\frac{\partial \ln p_{\mathbf{R}|A}(\mathbf{r}|a)}{\partial a} \right]^2 \right\} \right)^{-1} \tag{3.14}$$

or equivalently

$$\text{Var}\left[\hat{a}(\mathbf{R}) - a\right] \geq \left(-E\left[\frac{\partial^2 \ln p_{\mathbf{R}|A}(\mathbf{r}|a)}{\partial a^2} \right] \right)^{-1} \tag{3.15}$$

where the first and second partial derivatives are assumed to exist and be absolutely integrable. Any estimate that satisfies this bound is called an efficient estimate. The inverse of the bound is referred to as the Fisher information.

3.5 Multiple Parameter Estimation

All of the above techniques can be applied to multiple parameter estimation problems by simply forming a vector of parameters to be estimated. This section is dedicated to demonstrating this and to defining a few operators to simplify notation.

Let $\mathbf{a} = [a_1 \ a_2 \ a_3 \ .... \ a_i]^T$ be a vector of parameters that we wish to estimate. From this we can write a joint log-likelihood function

$$\mathcal{L}(\mathbf{a}) = \ln\left[p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}|\mathbf{a}) \right] \tag{3.16}$$

The maximum likelihood estimate is found by

$$\nabla_a \left[\mathcal{L}(\mathbf{a}) \right] \big|_{\mathbf{a}=\hat{\mathbf{a}}} \tag{3.17}$$

where

$$\nabla_a \equiv \left[\frac{\partial}{\partial a_1} \ \frac{\partial}{\partial a_2} \ \frac{\partial}{\partial a_3} \cdots \frac{\partial}{\partial a_i} \right]^T \tag{3.18}$$

3.5.1 Estimator Quality.

Bias. Since we have an estimate vector, we will also have a bias vector. The bias vector is defined as

$$\mathbf{B}(\mathbf{a}) \equiv E[\mathbf{a}_\epsilon(\mathbf{R})] \tag{3.19}$$

where

$$\mathbf{a}_\epsilon(\mathbf{R}) = \hat{\mathbf{a}}(\mathbf{R}) - \mathbf{a} \tag{3.20}$$

and the expected value of the vector is defined by

$$E[\mathbf{a}] = [E[a_1] \ E[a_2] \ E[a_3] \ ...E[a_N]]^T \tag{3.21}$$

We call an estimate unbiased if each component of the bias vector is zero.

Covariance Matrix. For the single parameter case we define the spread of the error by the variance of the estimate; for the multiple parameter case the analogous quantity is the covariance matrix

$$\mathbf{\Lambda}_\epsilon = E[(\mathbf{a}_\epsilon - \mathbf{B}(\mathbf{a}))(\mathbf{a}_\epsilon - \mathbf{B}(\mathbf{a}))^T] \tag{3.22}$$

Much like in the single parameter case, it is not always practical to try and calculate the values in $\mathbf{\Lambda}_\epsilon$. Instead we will calculate a lower bound. First we find the Fisher

information matrix, J, whose elements are found by

$$J_{ij} \equiv E \left[\frac{\partial \ln \left[p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}|\mathbf{a}) \right]}{\partial a_i} \cdot \frac{\partial \ln \left[p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}|\mathbf{a}) \right]}{\partial a_j} \right] \tag{3.23}$$

which can also be written as

$$J_{ij} \equiv -E \left[\frac{\partial^2 \ln \left[p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}|\mathbf{a}) \right]}{\partial a_i \partial a_j} \right] \tag{3.24}$$

Now let $\mathbf{K} = \mathbf{J}^{-1}$. The lower bounds on the variance of a_i are

$$\Lambda_{ii} \geq K_{ii} \tag{3.25}$$

IV. A Review of Deconvolution

4.1 Introduction

The purpose of this chapter is to report the results of work done by MacDonald [23]; the work is foundational to this research and is therefore worth including. Much of this chapter is adapted from his dissertation. This chapter will also demonstrate a modified stopping method for the algorithm that we believe to be more accurate and reliable.

4.2 Problem description and geometry

As stated previously, deconvolution is one method for recovering images using image plane data; a simple imaging system is shown in Figure 4.1. Most deconvolution methods are applicable to incoherent imaging, however this research deals with imaging coherently illuminated objects. A coherent speckled image cannot be directly processed using traditional deconvolution techniques. To use deconvolution an incoherent image must be formed (or approximated) from many frames of coherent imagery [25].

4.2.1 Geometry. The coordinate system for the deconvolution problem will use (x, y) to describe positions in the object plane and (u, v) to describe position in the image plane. To simplify notation, we will define the following variable in $\mathbb{R}^2$ to represent the ordered pairs
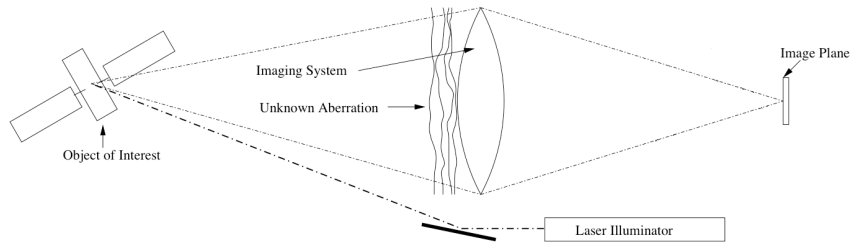
$$X = (x, y)$$

$$U = (u, v)$$

Figure 4.1:   Illustration of a simple imaging system

4.2.2   Data and statistical models.   In order to derive an estimator one must have information on the statistics of the data, as well as understand what the ideal image that would be formed in the absence of noise.

4.2.2.1   Data Model.   The model for the image plane irradiance takes advantage of the fact that the multi-frame average of laser speckle images converges to the incoherent model [24]. This is due to the assumption that the phase at the target is random and independent from observation to observation in a manner consistent with the time varying phase distribution produced by incoherent light. By capitalizing on this, the image plane model becomes a convolution of the geometric image of the source irradiance and a Point Spread Function (PSF). The PSF would include the effects of the optical system and the path turbulence. This convolution is written discretely to facilitate the derivation of a computer algorithm for image recovery.

$$i(U) = [o * s_i](U) = \sum_X o(X)s_i(U - X) \tag{4.1}$$

where $o(X)$ is the geometric image of the source intensity and $s_i(U)$ is the image plane PSF. Since in cases of interest to this research, the aperture is large with respect to the Fried parameter of the turbulence, the PSF is dominated by atmospheric effects. The long exposure PSF must be used since the Signal to Noise Ratio (SNR) of fully developed speckled frames is always equal to 1; making tip-tilt removal problematic.

To complete the data model, we must look at the dominate noise sources in the system, and form an appropriate statistical model to describe the corrupted data.

4.2.2.2   Statistical Model.   To facilitate forming a statistical model we will define a single frame of image plane data as

$$d_i^{(k)}(U) = i(U) + n_{shot}^{(k)}(U) + n_{speckle}^{(k)}(U) \tag{4.2}$$

33

where i is the ideal incoherent image, $n_{shot}^{(k)}$ is noise associated with random photon arrivals, and $n_{speckle}^{(k)}$ is a noise term that captures the effects of speckle and any other deviations from the model resulting from the coherent illumination. Both speckle and shot noise are assumed independent from pixel to pixel; this assumption allows for computational efficiency. The assumption for shot noise is common and will not be rigorously justified; however, for the case of speckle noise we will investigate the validity of this assumption in Appendix C. If we sum $\mathcal{M}$ frames of speckle image data, d_i^k, we have a new random variable, d_i. We are interested in describing the statistics of d_i. When $\mathcal{M} = 1$ the dominate noise is $n_{speckle}$ and d_i is exponential distributed. As $\mathcal{M} \rightarrow \infty$ then n_{shot} becomes the dominate noise source and d_i is Poisson distributed. It would be desirable to have a single distribution that was valid for all $\mathcal{M}$. The distribution that best captures all of these effects is the negative binomial distribution [19]

$$p(K) = \frac{\Gamma(K + \mathcal{M})}{\Gamma(K + 1)\Gamma(\mathcal{M})} \left[1 + \frac{\mathcal{M}}{\overline{K}}\right]^{-K} \left[1 + \frac{\overline{K}}{\mathcal{M}}\right]^{-\mathcal{M}} \tag{4.3}$$

where K is a random variable representing photocounts and $\overline{K}$ is the mean number of photocounts.

4.3 Derivation

Start with the likelihood function based on the negative binomial distribution [23]

$$\mathcal{L}(o) = \sum_U d_i(U) \ln[i(U)] - [d_i(U) + \mathcal{M}] \ln[i(U) + \mathcal{M}] \tag{4.4}$$

Since we are summing multiple frames of coherent data, the likelihood function of partially coherent data is used, where $\mathcal{M}$ is the number of frames. The model for $i(U)$ is the incoherent model, which is easily described as a convolution

$$i(U) = \sum_X o(X)s_i(U - X) \tag{4.5}$$

34

Using this definition of $i(U)$ we can write $\mathcal{L}(o)$ to show the dependance on o

$$\mathcal{L}(o) = \sum_U d_i(U) \ln[\sum_X o(X)h(U-X)] - [d_i(U) + \mathcal{M}] \ln[\sum_X o(X)h(U-X) + \mathcal{M}] \quad (4.6)$$

Recall that a maximum likelihood estimate is defined by

$$\frac{d\mathcal{L}(o)}{do}\big|_{o=\hat{o}} = 0 \quad (4.7)$$

Solving for the required derivative we arrive at

$$\frac{d\mathcal{L}(o)}{do} = \sum_U \frac{d_i(U)}{i(U)} h(U-X) - \frac{d_i(U) + \mathcal{M}}{i(U) + \mathcal{M}} h(U-X) \quad (4.8)$$

Next we implement a Richardson-Lucy type algorithm to solve for o

$$o_{new}(X) = o_{old}(X) \frac{\sum_U \frac{d_i(U)}{i(U)} h(U-X)}{\sum_U \frac{d_i(U) + \mathcal{M}}{i(U) + \mathcal{M}} h(U-X)} \quad (4.9)$$

Implementation is simplified by recognizing that this can be written in terms of correlations

$$o_{new}(X) = o_{old}(X) \frac{[\frac{d_i}{i} \star h](X)}{[\frac{d_i + \mathcal{M}}{i + \mathcal{M}} \star h](X)} \quad (4.10)$$

where $\star$ is used to represent the correlation operation.

4.4 Stopping the algorithm

In order for this algorithm to be useful, we have to be able to stop it at or near the optimal iteration. MacDonald demonstrated one technique [23], but it relied on some assumptions that we would like to remove. Any stopping criteria should be based on the statistics of the data and not require *a priori* knowledge of the object. Since we cannot reliably use estimation to remove the noise, the algorithm should be stopped when the variance of the data, d_i, around the estimated mean, i, are close to

the variance of the data around its sample mean

$$K^{-1} \sum_{k=1}^{K} (d_i^k - \hat{i})^2 < \alpha \sigma_i^2 \qquad (4.11)$$

where $\hat{i}$ is the image estimate, σ_i^2 is the variance of the data around the sample mean, and α is a parameter that lets the user choose the degree of damping. This parameter will be discussed later. When the above condition is satisfied for a given pixel, the iterations for that pixel are terminated. In order to stop each pixel independently we add a binary mask,m_i, to our algorithm

$$\hat{o} = \hat{o}_{old} \frac{\left[\left[(1 - m_i)\frac{d_i}{\hat{i}} + m_i \right] \star s_i \right]}{\left[\left[(1 - m_i)\frac{d_i + \mathcal{M}}{\hat{i} + \mathcal{M}} + m_i \right] \star s_i \right]} \qquad (4.12)$$

When the algorithm begins m_i, is a matrix of zeros; the algorithm is then checked at each iteration to find the pixels that satisfy Equation 4.11. When a pixel its found that satisfies the damping criteria m_i is set to one for that pixel. This effectively stops the iteration at that pixel by forcing $d_i(U_0) = \hat{i}(U_0)$.

The damping parameter, α, gives the user flexibility to choose the level of damping. When α is chosen too low, it is possible to over-iterate and amplify noise; if α is chosen too high the iteration will slow considerably. In order to provide some guidance on choosing α, we turn to the *strong law of large numbers* [6] which states that for independent, identically distributed random variables the sample mean will approach the true mean, with probability 1, as the number of sample increases. For our purposes this means α is inversely proportional to the number of frames; this relationship is assumed to be linear for simplicity. The relationship is not truly linear, but it is believed a simple relationship will work over the range of frame numbers for this research. Using this as a starting point, an attempt was made to empirically determine a function to give an "ideal" value for α.

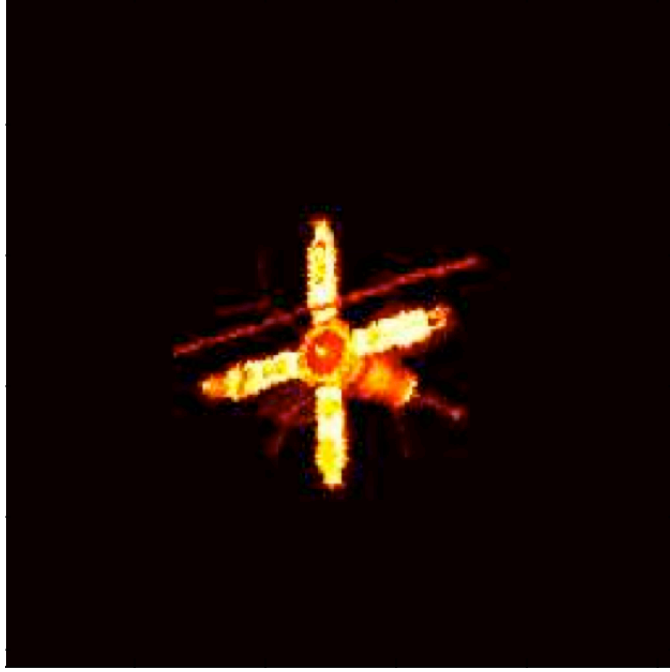$$\alpha = 1 + \frac{\hat{\alpha}}{\mathcal{M}} \qquad (4.13)$$

Figure 4.2:   Satellite image used as the truth data in simulated data sets

where $\hat{\alpha}$ will be optimized experimentally.

4.5   Data Simulation

The simulated data sets were created using Matlab® . Each set consists of 200 independently created frames of both pupil plane and image plane data. In order to create the frames of data a field magnitude was defined in the object plane; the magnitude in the object plane is shown in Figure 4.2. A random phase (uniform on $(-n\pi, n\pi]$) was added to every point in the field to model the surface roughness. Each phase is constructed to be independent of every other phase in the field. The target was sampled at the Nyquist rate required by the pupil plane aperture. The sample rate, Δ, is calculated according to

$$\Delta = \frac{\lambda z}{2D} \tag{4.14}$$

where λ is the optical wavelength, z is the propagation distance, and D is the diameter of the aperture. This field is then propagated to the pupil plane using a Fourier

transform. It can be shown that due to the random phase in the target field any propagation, regardless of propagation length, can be modeled as a Fraunhoffer propagation and therefore a simple Fourier transform. The imaging aperture function and the turbulence phase screen are then applied to the pupil plane field and an inverse Fourier transform was performed to give us the image plane field; the magnitude squared gives us the image plane intensity. In equation form this would be:

$$i(U) = \left| \mathcal{F}^{-1} \left\{ \mathcal{F} \left\{ U(X) \right\} P \exp\{j\theta\} \right\} \right|^2 \tag{4.15}$$

where P is the aperture function and θ is the phase screen.

To account for detection noise processes, both Gaussian-distributed read noise and Poisson distributed shot noise were applied to the data. Gaussian detector noise was characterized by the standard deviation σ_d. The average detection SNR was then defined by

$$\text{SNR} = \frac{p}{\sqrt{p + \sigma_d^2}} \tag{4.16}$$

where p is the average number of photo-electrons per pixel in the data set. To generate data with a desired detection SNR, Equation 4.16 was inverted and the data scaled to the required average p. The scaled data was then used to generate Poisson-distributed random numbers with mean p. Uncorrelated Gaussian random numbers with a standard deviation of σ_d were then added to each pixel. SNR was defined in this manner to be consistent with earlier work [22].

4.6   Results and Conclusions

With the deconvolution algorithm derived and implemented, the next step is to characterize its performance. The approach taken for performance characterization was to use a Monte-Carlo simulation consisting of 100 data sets of 200 frames each at four different turbulence levels. It is often difficult to chose metrics to quantify the quality of a recovered image, since for almost any metric an object estimate can be chosen that is much better or worse than the chosen metric implies. Initially mean

absolute error (MAE) was chosen as a metric. MAE is defined as

$$MAE = (NM)^{-1} \sum_{n=1,m=1}^{N,M} |o - \hat{o}| \tag{4.17}$$

This metric was quickly discarded as the errors were excessively large for estimates that were shifted with respect to the truth object. To avoid this problem we choose a slightly more complicated metric that is translation invariant

$$Er = \min_{u_0,v_0} \frac{\sum |\hat{o}(u - u_0, v - v_0) - o(u, v)|^2}{\sum |o(u, v)|^2} \tag{4.18}$$

which is more easily calculated as

$$Er = \frac{r_{\hat{o}\hat{o}}(0, 0) + r_{oo}(0, 0) - 2 \max_{u_0,v_0} Re\{r_{o\hat{o}}(u_0, v_0)\}}{r_{oo}(0, 0)} \tag{4.19}$$

where r_{oo} and $r_{\hat{o}\hat{o}}$ are autocorrelations of the object and the estimate, and $r_{o\hat{o}}$ is the cross correlation. In general correlation function are defined as

$$r_{ab}(u, v) = \sum_x \sum_y a(x, y)b(u + a, v + b) \tag{4.20}$$

For a more complete explanation see [9]. An estimator will seek to minimize this metric, Er.

The first step in characterization was to find an optimal value for the damping parameter; this was done using a single turbulence value and then assumed to be correct at any turbulence level; $\hat{\alpha} = 15$ was chosen. The damping parameter was chosen by running the simulation at various values of $\hat{\alpha}$ and choosing the best value. Once the damping parameter was determined, the image recovery algorithm was run on all the data sets and the results are presented below.

4.6.1 Baseline results. In order to determine the effects that certain system parameters (frames, turbulence) have on system performance we need to define and
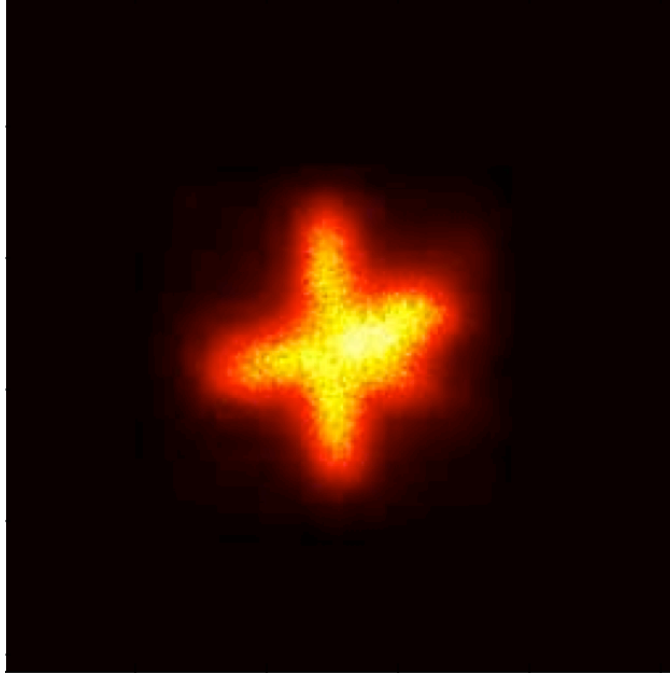
Figure 4.3:   Sum of 200 frames of raw image plane data

characterize a baseline case. For this research the baseline case is a 200 frame data
set generated in turbulence described by $\frac{D}{r_0} = 10$. The sum of one set (200 frames)
of raw image plane data is shown in Figure 4.3. A sample image recovered from this
scenario can be seen in Figure 4.4 Clearly the recovered image is improved over the
raw data, however no high spatial frequency details have been recovered. We have
a better idea of the shape and extent of the object, but little information about the
fine structure. To quantify the improvement, the error function (Er) was calculated
at each iteration. This was done for 100 data sets and Er was plotted with error bars
representing +/- one standard deviation. The results are shown in Figure 4.5

4.6.2   Effects of varying the number of data frames.   Recalling from the ear-
lier development that the model for the data was based on the incoherent convolution
model, it should stand to reason that as the number of frames increases the data will
more closely match the model and therefore the recovered images should be better.
The effects of the number of frames will be quantified by running the deconvolution
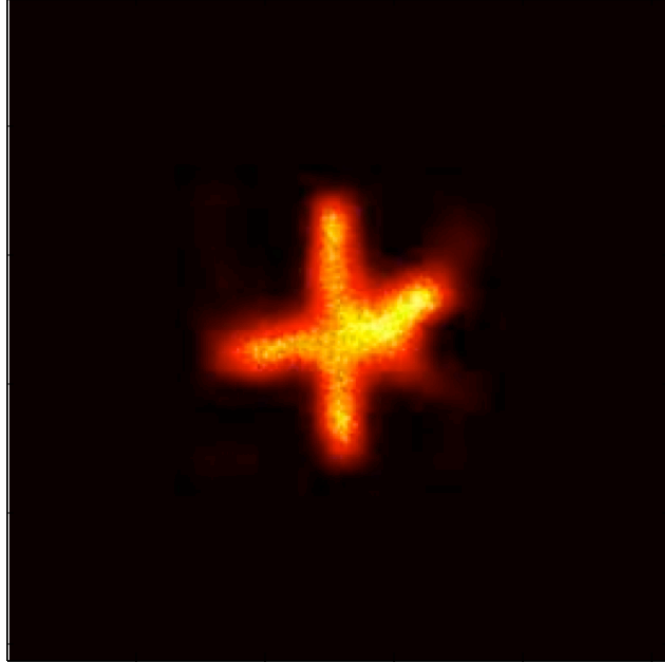algorithm on data sets of 50 and 100 frames and comparing the recoveries to the

40

Figure 4.4:   Sample image recovered using 3500 iterations of deconvolution
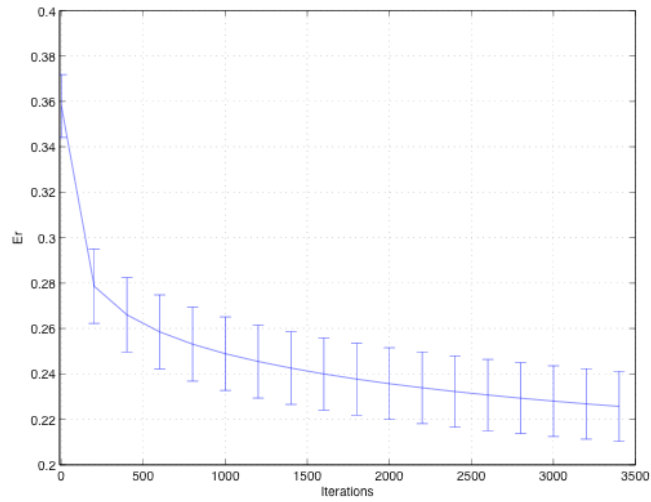

Figure 4.5: Er versus iterations for baseline simulation. The mean of 100 recoveries was plotted with error bars representing +/- one standard deviation

200 frame reconstructions. Figure 4.6 shows images recovered using 50, 100 and 200 frames of data; while the differences are not dramatic, clearly the quality of the image improves with additional frames. To really understand the benefit of increasing the number of data frames we look at the variance of Er for the object estimates. As seen in Figure 4.7, the variance increases dramatically as the number of frames decrease. What this means from a practical viewpoint is that as the number of frames increases so does the reliability of the algorithm.

4.6.3   Effects of turbulence strength on the recovered image.   While it should be obvious to the informed reader that the strength of the atmospheric turbulence will have a direct impact on the quality of the recovered image, it will still be quantified here. It is shown to illustrate that the joint algorithm (see Chapter 6) will be less affected by turbulence. Figure 7.2 shows Er for the recovered image at each iteration for 4 different turbulence strengths. As expected turbulence has a large impact on the quality of the reconstruction.

4.6.4   Ability to recover varied intensity images.   This section seeks to quantify the ability of the deconvolution algorithm to correctly recover the various intensities in a single image accurately. A truth image consisting of five squares of varying intensity is shown in Figure 4.9; the center square has an intensity of 1, while the other four have intensities of 0.1, 0.25, 0.5, and 0.75 (starting in the upper left and moving clockwise). In the recovered image it is desired for the ratios of intensity in each corner to the intensity of the center square to be accurately recovered. The recovered image can be seen in Figure 4.10.   The mean intensities in each corner square (as a percentage of the mean intensity in the center square) in the recovered image are 9.78%, 23.7%, 48.8% and 73%. Clearly the algorithm can recover the proper intensity values in gray scale images. It is further noted that greater than 90% of the energy in the field of view is in the image.
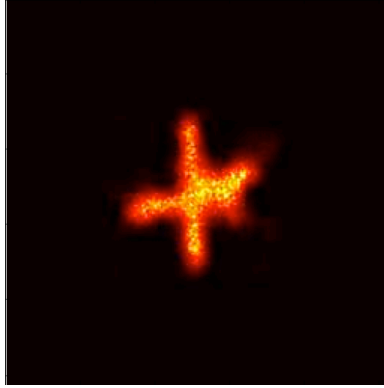
(a) 50 Frames
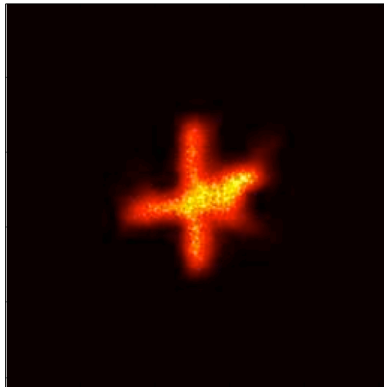
(b) 100 Frames
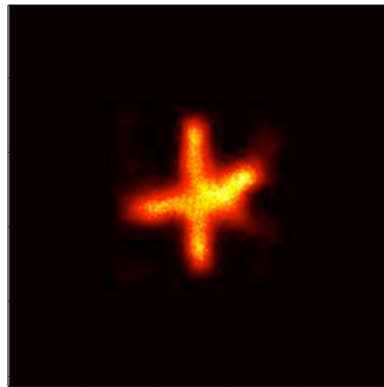
(c) 200 Frames

Figure 4.6:   Images recovered using different numbers of data frames.
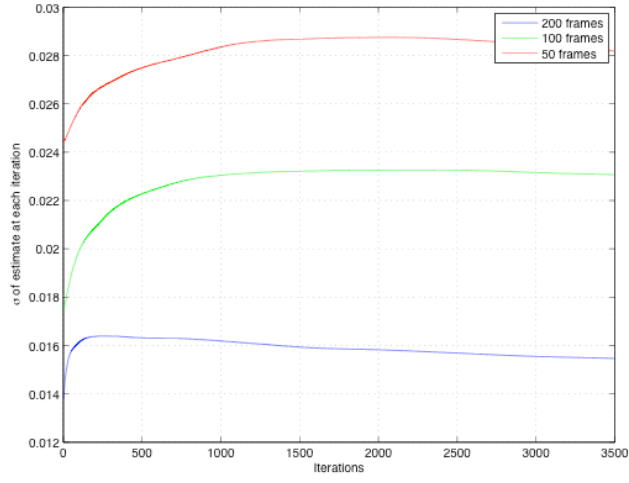
Figure 4.7:   Variance of the estimate for varying numbers of frames of data
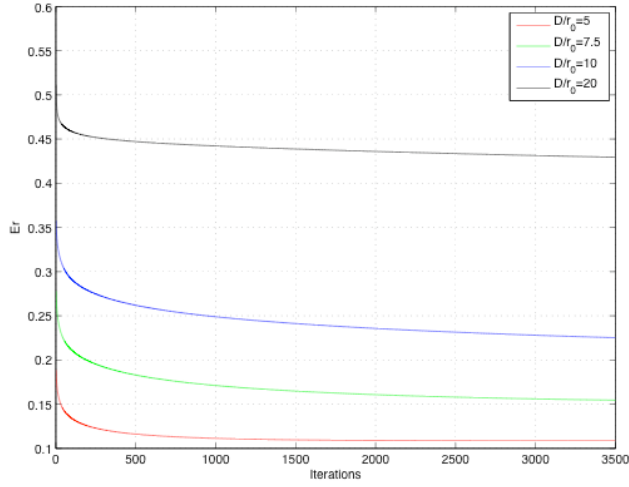

Figure 4.8: Er of recovered images shown for varying turbulence levels
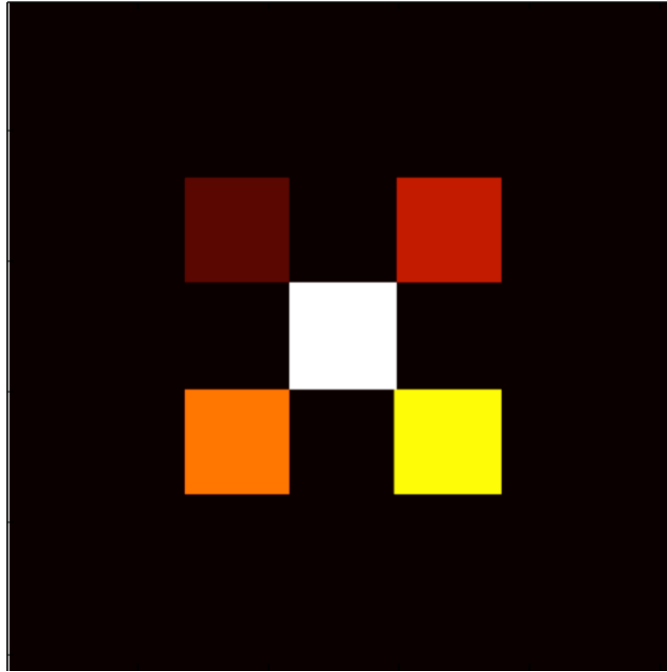
44

Figure 4.9:   Truth image used for quantifying intensity recovery accuracy
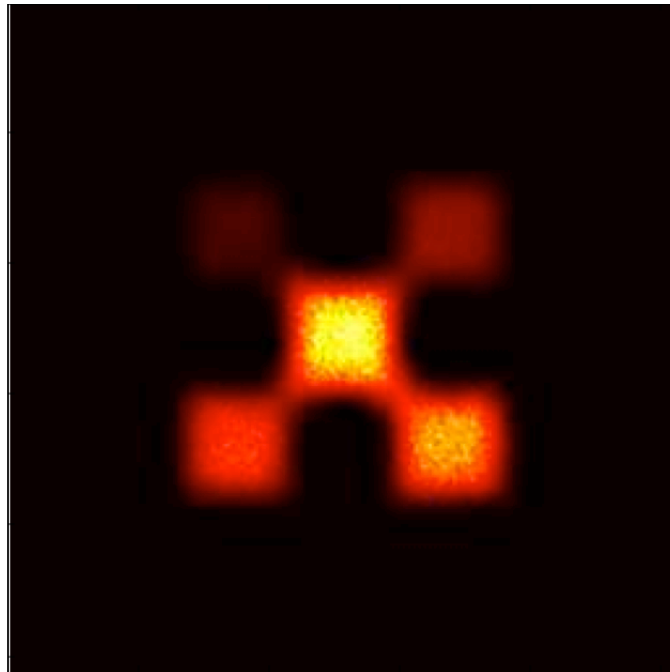

Figure 4.10:   Recovered gray scale image.

4.6.5   Quantification of resolution of the algorithm.   To define the resolution of the algorithm, various bar targets are imaging. The system resolution is defined as the pixel spacing between bars in the targets when the algorithm can just resolve them; the bars will be said to be resolved when the intensity of the valleys fall to less than half that of the peaks, and any additional bars in the recovered image are suppressed by at least a factor of 5. The deconvolution algorithm was able to resolve the bars when they were separated by 10 pixels. The results are shown in Figurea s4.11-4.12.   One could argue that the human eye can separate the bars in the raw data, however the criteria for resolution is the valleys must have less than half the photons of the peaks. This is clearly not the case for the raw data, but is satisfied by the deconvolved case. The slices that are plotted in Figures 4.11b) and (4.12b are the mean of 20 vertical slices from the center of the image data.
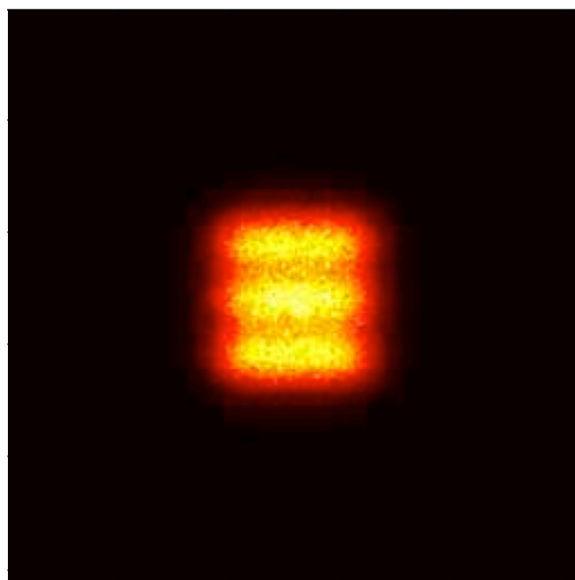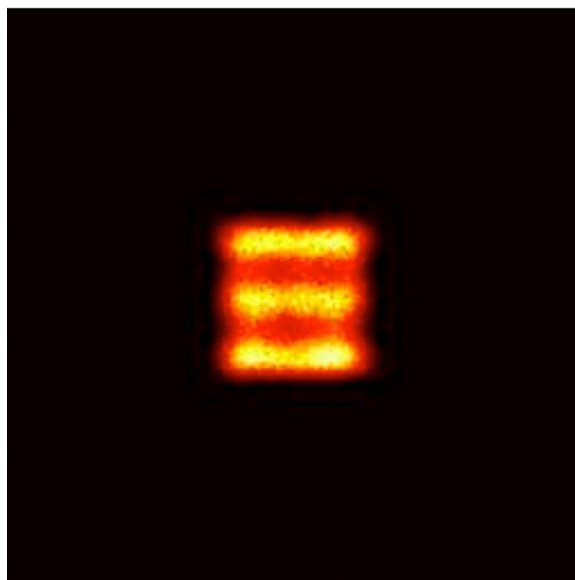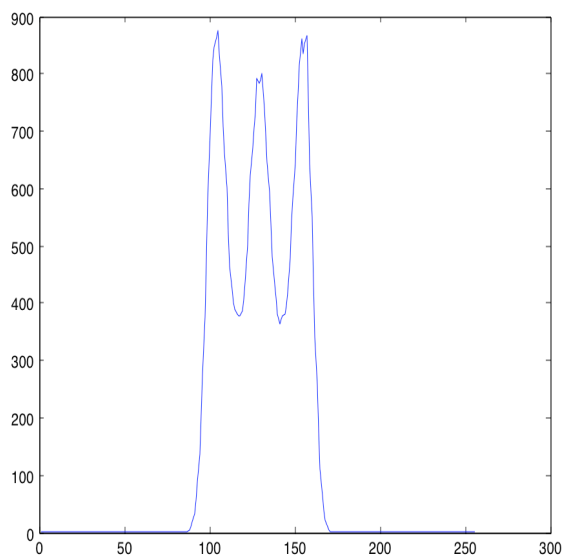
(a)
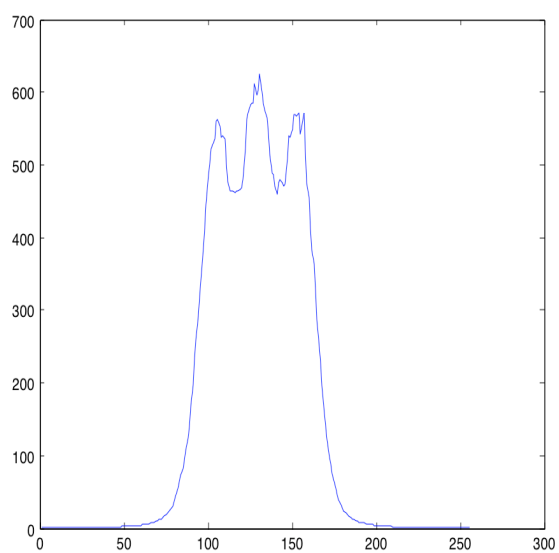
(b)

Figure 4.11:   Raw image data of a resolution target. 10 pixels separate the bars

(a)

(b)

Figure 4.12:   Deconvolved image data of a resolution target. 10 pixels separate the bars

V. Maximum Likelihood Pupil Plane Image Recovery

5.1   Introduction

This chapter will demonstrate a high resolution pupil plane imaging system. The algorithm proposed here is a Maximum Likelihood pupil plane algorithm.

5.2   Problem description and geometry

We have previously described methods for recovering images from pupil plane intensity measurements. This can have many advantages since the arrays can be made large and conformal. This raises the theoretical resolution limit without a large increase in system volume. However, pupil plane systems suffer from some widely known stagnation problems [15]. This chapter derives and implements a new pupil plane algorithm. This new algorithm does not eliminate the existing stagnation issues, but is formed in a manner that will allow it to be easily fused with the image plane system of chapter 4.

5.2.1   Geometry and Assumptions.   The coordinate system for all the following derivations will use (x, y) functions for positions in the object plane, (α, β) in the pupil plane, (u, v) in the image plane, and (ξ, η) for shifts in the autocorrelations. However to simplify notation, we define the following variables in $\mathbb{R}^2$ to represent the ordered pairs

$$X = (x, y)$$

$$\Lambda = (\alpha, \beta)$$

$$\Psi = (\xi, \eta)$$

$$U = (u, v)$$

It also must be pointed out that for the entire paper any matrix product is a Hadamard (or element wise) product.

5.2.2 Pupil Plane Data Model. The pupil plane model is based on work from Fienup and Idell [21], which relates the autocorrelation of the object brightness function to the average magnitude squared of the Fourier transform of the pupil plane intensity. The autocorrelation of the object brightness function is given by

$$R_0(\Psi) = \sum_X o(X)o(\Psi + X) \tag{5.1}$$

Image recovery begins with estimating the average energy spectrum of the observed speckle pattern by averaging the squared moduli of many independent speckled autocorrelations

$$R_0^{(k)}(\Psi) = \mathcal{F}_{\Lambda,\Psi}^{-1} \left[|\mathcal{F}_{X,\Lambda}^{\lambda z}[f^{(k)}(X)]|^2 H(\Lambda) \right] \tag{5.2}$$

where $f^{(k)}$ is the kth realization of the field reflected from the object and the aperture function, H defines the region over which the speckle pattern is observed in the pupil plane, $\mathcal{F}_{\Lambda,\Psi}^{-1}$ in an inverse Fourier transform which operates in a function in Λ and returns a function of Ψ, and $\mathcal{F}_{X,\Lambda}^{\lambda z}$ is a Fraunhoffer propagation operator [20]. It has been shown that as the number of independent speckle realizations, J, increases the average energy spectrum converges to [21]

$$\lim_{K \to \infty} K^{-1} \sum_{k=1}^{K} |R_0^{(k)}(\Psi)|^2 = bs_p(\Psi) + cR_0(\Psi) * s_p(\Psi) \tag{5.3}$$

where b and c are constants, and $s_p(\Psi) = |h(\Psi)|^2$ is the PSF of the pupil plane aperture.

The intensity in the pupil plane is related to the object field by

$$pupil^{(k)}(\Lambda) = |\mathcal{F}_{X,\Lambda}^{\lambda z}\{f^k(X)\}|^2 + n_{shot}^{(k)}(\Lambda) + n_{speckle}^{(k)}(\Lambda) \tag{5.4}$$

where $n_{shot}^{(k)}$ is shot noise from random photon arrival times and $n_{speckle}^{(k)}$ is the speckle noise associated with coherent systems. It is convenient at this point to define a

variable to represent a single frame of transformed pupil plane data

$$d_r^{(k)}(\Psi) = \left| \mathcal{F}_{\Lambda,\Psi}^{-1} \left\{ |\mathcal{F}_{X,\Lambda}^{\lambda z}\{f^{(k)}(X)\}|^2 + n_{shot}^{(k)}(\Lambda) + n_{speckle}^{(k)}(\Lambda) \right\} \right|^2 \tag{5.5}$$

where k is the frame number. According to Equation 5.2, we can average many frames of this transformed data to represent the average speckled autocorrelation of the object brightness function.

$$d_r(\Psi) = K^{-1} \sum_{k=1}^{K} d_r^{(k)}(\Psi) \tag{5.6}$$

Provided that K is large d_r will be approximately Gaussian with mean $R(\Psi) = b|h(\Psi)|^2 + cR_0(\Psi) * |h(\Psi)|^2$. From this we can write an equation for the probability distribution of a single pixel

$$p_{D_r(\Psi)}(d_r(\Psi)) = \frac{exp\left\{\frac{-[d_r(\Psi)-R(\Psi)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi}\sigma} \tag{5.7}$$

If we also assume that the noise in each pixel is independent of the noise in all the other pixels we can write the equation for the probability of realizing an entire "scene"

$$p_{D_r}(d_r|o,h) = \prod_{\Psi} \frac{exp\left\{\frac{-[d_r(\Psi)-R(\Psi)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi}\sigma} \tag{5.8}$$

5.3 Image Recovery Algorithm

The algorithm development begins by defining the log-likelihood function

$$\mathcal{L}(o,b,c) = \ln p_{D_r}(d_r|o,s_p) = \sum_{\Psi} \frac{-[d_r(\Psi) - R(\Psi,b,c)]^2}{2\sigma^2(\Psi)} \tag{5.9}$$

where

$$R(\Psi,b,c) = bs_p(\Psi) + c[R_0 * s_p](\Psi) \tag{5.10}$$

.

51

The maximum likelihood (ML) estimator is defined by [34]

$$\nabla \mathcal{L}(A)|_{A=\hat{a}} = 0 \tag{5.11}$$

where $A = [o\ b\ c]^T$ and $\hat{a} = [\hat{o}\ \hat{b}\ \hat{c}]^T$

An estimate of o is difficult to find in closed form, but rather can be solved iteratively after we have initial estimates of b and c.

5.3.1 Estimates of b and c. To solve for initial estimates of b and c we restate the problem as the likelihood function of b and c conditioned on o

$$\mathcal{L}(b, c|o) = \sum_{\Psi} \frac{-[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)]^2}{2\sigma^2(\Psi)} \tag{5.12}$$

The new estimation routine is defined by

$$\nabla \mathcal{L}(A)|_{A=\hat{a}} = 0 \tag{5.13}$$

where $A = [b\ c]^T$ and $\hat{a} = [\hat{b}\ \hat{c}]^T$. These estimates will be updated as the value of o is refined. To find values of $\hat{b}$ and $\hat{c}$ we must find the gradient of $\mathcal{L}(A)$

$$\nabla \mathcal{L}(b, c) = \left[\frac{\partial \mathcal{L}(b, c)}{\partial b} \quad \frac{\partial \mathcal{L}(b, c)}{\partial c} \right]^T \tag{5.14}$$

Solving these two derivatives separately and setting them equal to zero gives us a set of simultaneous equations.

$$\frac{\partial \mathcal{L}(b, c)}{\partial b} = \sum_{\Psi} \frac{[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)]s_p(\Psi)}{\sigma^2(\Psi)} = 0 \tag{5.15}$$

$$\frac{\partial \mathcal{L}(b, c)}{\partial c} = \sum_{\Psi} \frac{[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)][R_0 * s_p](\Psi)}{\sigma^2(\Psi)} = 0 \tag{5.16}$$

Next we solve Equation 6.22 for c as a function of b

$$c = \frac{\sum_{\Psi}[d_r(\Psi) - bs_p(\Psi)][R_0 * s_p](\Psi)}{\sum_{\Psi}[R * s_p]^2(\Psi)} \tag{5.17}$$

Substitute Equation 6.23 into Equation 6.21 and solve for b

$$\hat{b} = \frac{\sum_{\Psi} d_r(\Psi)s_p(\Psi) - \rho_1 \sum_{\Psi} s_p(\Psi)[R_0 * s_p](\Psi)}{\sum_{\Psi} s_p^2(\Psi) - \rho_2 \sum_{\Psi} s_p(\Psi)[R_0 * s_p](\Psi)} \tag{5.18}$$

where

$$\rho_1 = \frac{\sum_{\Psi'} d_r(\Psi')[R_0 * s_p](\Psi')}{\sum_{\Psi'}[R_0 * s_p]^2(\Psi')} \tag{5.19}$$

and

$$\rho_2 = \frac{\sum_{\Psi'} h(\Psi')[R_0 * s_p](\Psi')}{\sum_{\Psi'}[R_0 * s_p]^2(\Psi')} \tag{5.20}$$

Finally, substitute the value of $\hat{b}$ into Equation 6.23 to calculate a value for $\hat{c}$.

5.3.2 Estimation algorithm for o. Next we will build an iterative algorithm to solve for o. It is informative here to write the log-likelihood function to show the dependance on o. We do this by recognizing

$$[R_0 * s_p](\Psi) = \sum_U \sum_X o(X)o(U + X)s_p(\Psi - U) \tag{5.21}$$

Combining this result with Equation 5.10 and (6.14) we are able to write

$$\mathcal{L}(o|b,c) = \sum_{\Psi} \frac{-[d_r(\Psi) - bs_p(\Psi) - c\sum_U \sum_X o(X)o(U + X)s_p(\Psi - U)(\Psi)]^2}{2\sigma^2(\Psi)} \tag{5.22}$$

Next we take the derivative of this log-likelihood with respect to o

$$\frac{\partial \mathcal{L}(o|b,c)}{\partial o} = \sum_{\Psi} \frac{[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)]c\sum_U \frac{\partial R_0}{\partial o}s_p(\Psi - U)}{\sigma^2(\Psi)}$$

where

$$\frac{\partial R}{\partial o} = o(X + U) + o(X - U) \tag{5.23}$$

53

Below are a few definitions to simplify this equation

$$c \sum_U \frac{\partial R_0}{\partial o} s_p(\Psi - U) = c\{[o * s_p](\Psi + X) + [o \star s_p](\Psi - X)\} \qquad (5.24)$$

$$
\begin{aligned}
\Phi_1(\Psi + X) &= [o * s_p](\Psi + X) \\
\Phi_2(\Psi - X) &= [o \star s_p](\Psi - X)
\end{aligned} \qquad (5.25)
$$

$$\rho(\Psi) = \frac{b s_p(\Psi) + c[R_0 * s_p](\Psi)}{\sigma^2(\Psi)} \qquad (5.26)$$

Using the above definitions (and some further simplification) we are able to write a R-L algorithm

$$o_{new}(X) = o_{old}(X) \frac{c \left[\frac{d_r}{\sigma^2} \star \Phi_1\right](X) + c \left[\frac{d_r}{\sigma^2} * \Phi_2\right](X)}{[\rho \star \Phi_1](X) + [\rho * \Phi_2](X)} \qquad (5.27)$$

5.3.3 Stopping the Algorithm. Image recovery techniques built around a Richardson-Lucy algorithm are often ran for a set number of iterations and then exited. It would be better to have an optimized method of exiting the iterations. It is fairly well known that iterating beyond an optimal point can lead to noise amplification. This is due to the fact that all Maximum likelihood techniques attempt to fit the data as closely as possible given the constraints of the problem. This leads to the question of when do we stop our algorithm. Rather than stop the algorithm we will look at methods to damp the iteration to avoid over iteration regardless of how long the algorithm is run. This means one can iterate for as long as time allows with some assurance that iteration $z + 1$ will never be worse that iteration z.

The damping routine looks at the statistics of the data set and compares it to the model's statistics. When the variance of the model is near the variance of the

data that pixel is damped. Stated mathematically

$$K^{-1} \sum_{k=1}^{K} (d_r^k - \hat{R})^2 < \alpha \sigma_r^2 \tag{5.28}$$

where σ_r^2 is the measured variance of the data set and α is a user chosen value to determine the degree of damping (guidance for choosing α can be found in section(4.4)); $\hat{R}$ is defined as

$$\hat{R}(\Psi) = \sum_X o_{new}(X) o_{new}(\Psi + X) \tag{5.29}$$

Using the above criteria we can create a binary map, m_r, that will damp the iteration for the pixels where the criteria are satisfied. The values of the binary map will be set 1 for every pixel where the criteria are satisfied, otherwise the value is zero. The maps are updated at each iteration. The maps are applied to the update equation so that where ever the map equals 1 the data and the model are forced to be equal, essentially stopping that pixel from changing for that iteration. Using these criteria we restate Equation 5.26 and 5.27

$$\rho(\Psi) = bs_p(\Psi) + c \left[(1 - m_r)[R_0 * s_p](\Psi) + m_r \left[\frac{d_r - bs_p}{c} \right] \right] \tag{5.30}$$

$$o_{new}(X) = o_{old}(X) \frac{c \left[d_r \star \Phi_1 \right] (X) + c \left[d_r * \Phi_2 \right] (X)}{[\rho \star \Phi_1] (X) + [\rho * \Phi_2] (X)} \tag{5.31}$$

where m_r is the binary map applied to the pupil data.

5.4 Results

This section will compare the results of the proposed joint image recovery algorithm to that of deconvolution. This simulated data sets were created in a manner identical to that outlined in Chapter 4. After the data sets were created, two image recovery algorithms were run: 1) deconvolution [25] and 2)the pupil plane algorithm. The pupil plane algorithm is an intermediate step to a more complete joint algorithm that will be discussed in later chapters; for this reason the pupil plane analysis is not

as thorough as the deconvolution. The analysis will only consist of evaluating the resolution of the algorithm and the ability of the algorithm to recover varied intensity images. A section will also be devoted to showing some of the limitations of the algorithm as motivation for the remaining work.

5.4.1 Quantifying resolution.   The simulation was run in the absence of turbulence to get a diffraction limited image which is shown in Figure 5.1a. Clearly in the absence of turbulence the imaging system would have no problem resolving the bars in the target. The simulation is then run with $D/r_0 = 10$ and the resulting image is shown in Figure 5.1b. In this case none of the bars are resolved. In order to try and regain the lose resolution we apply the deconvolution algorithm and let it run to convergence; the resulting image is shown in Figure 5.1c and again the bars cannot be resolved. Figures 5.1d-e show the results of the pupil algorithm; clearly we have satisfied the resolution criteria from chapter 4. In chaper 4, it was shown that deconvolution had a resolution limit of 10 pixels; the pupil plane algorithm has a resolution limit of 5 pixels for the given turbulence scenario.

5.4.2 Ability to recover varied intensity images.   In section(4.6.4) it was shown that deconvolution algorithms are able to accurately recover various intensities in a single image. That same analysis was performed for the pupil plane algorithm with very different results. The pupil plane algorithm results are shown in Figure 5.2. Figure 5.2a shows the truth image and the recovered image is shown in Figure 5.2b. Clearly the pupil algorithm cannot recover the varied intensity in the image. This is easily explained by looking at Figure 5.2c, which shows the autocorrelation of the truth object. Since the pupil algorithm estimates the image from a corrupted measurement of the autocorrelation (which is symmetric) and the autocorrelation to object relationship is not one to one, the algorithm will choose an object that satisfies the known constraints. In the absence of information to the contrary, the algorithm will recover a symmetric object.
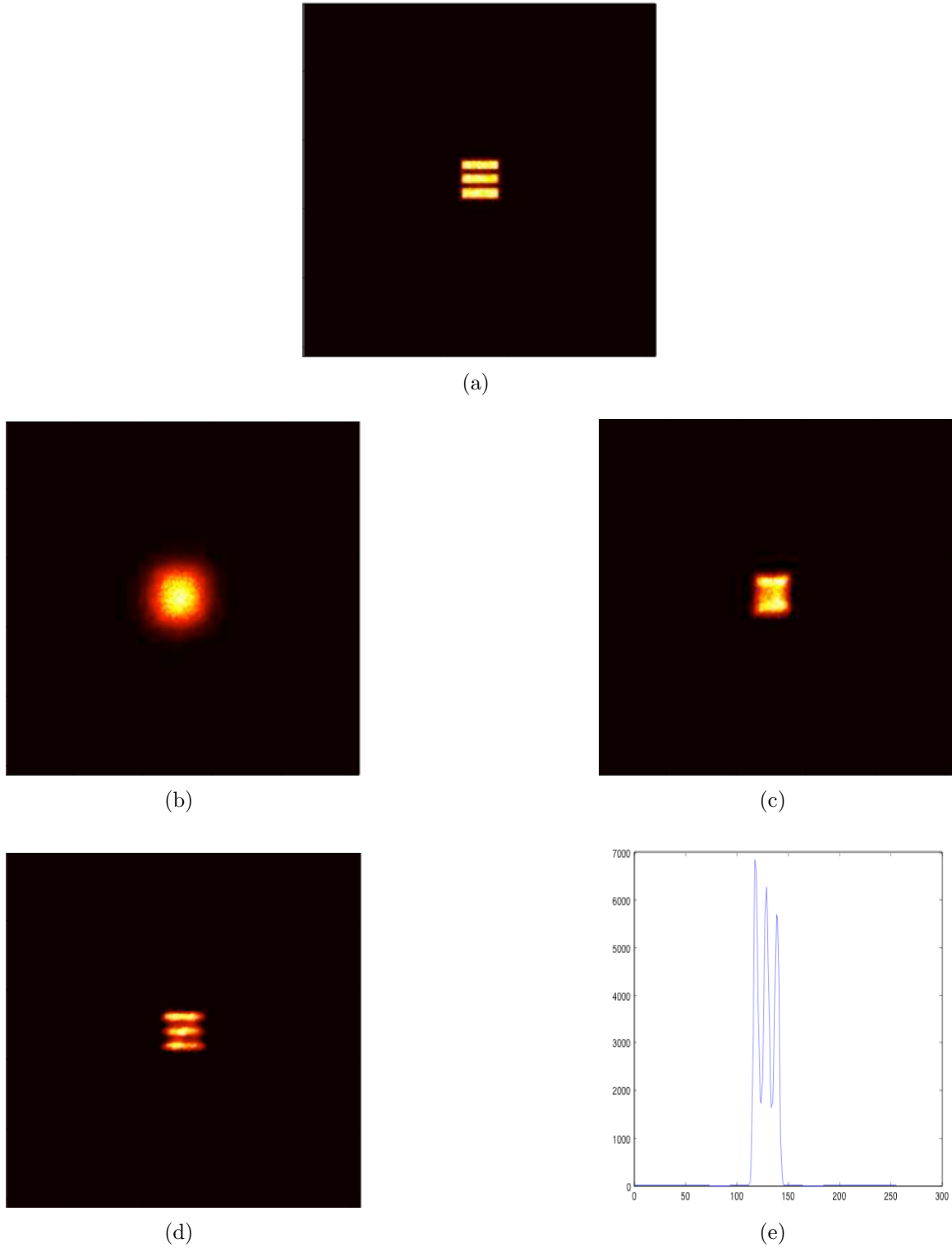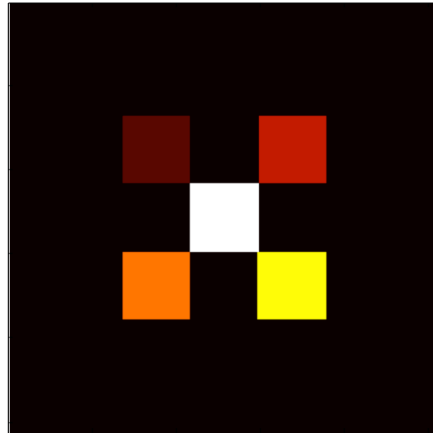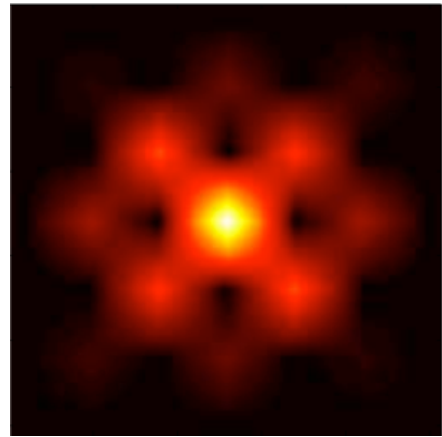
(a)

(b)

(c)

(d)

(e)

Figure 5.1:   Results of image recovery algorithms on a bar target where 5 pixels separate the bars. (a) show a diffraction limited image (b) show the raw data in the presence of turbulence (c) shows the results of deconvolution (d) shows the results of the pupil algorithm and (e) shows a slice throught (d)

57

(a)

(b)

(c)

Figure 5.2: (a) Image with intensity variations (b) Image recovered using pupil algorithm (c) autocorrelation of the truth object

5.4.3 Limitations of pupil plane algorithms.   The primary limitation to be discussed here is the inability of pupil plane algorithms to recover complex targets in the absence of good object support. The reason for this stems from the fact that the algorithms are generally based on using the object autocorrelation in the data model. Since an autocorrelation function is always symmetric and does not possess a one to one mapping with the underlying object, algorithms seeking to recover images from autocorrelations will have severe twin image stagnation problems. This can been seen in Figure 5.3; the satellite body is 'recovered' in both the lower right and the upper left of the satellite, while the smaller structures (antennas) are lost entirely.
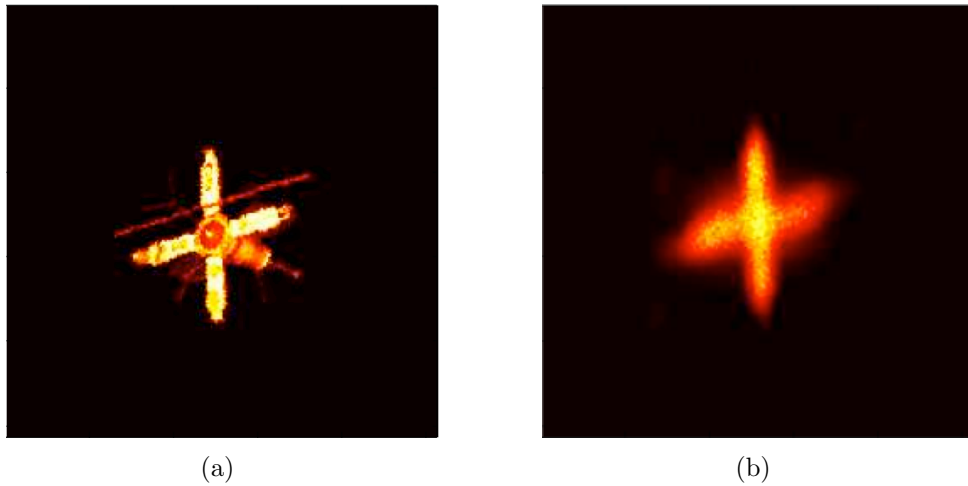

(a)   (b)

Figure 5.3:

VI. Joint Data Algorithm

6.1   Introduction

The chapter documents the fusion of the algorithms listed in the prior chapters. The algorithms are fused using Bayesian methods. The algorithm proposed here will use an image plane data set in conjunction with a pupil plane data set. From the two data sets a Maximum likelihood estimator will be formed. The use of image data with correlography methods is not new. Fienup suggested the use of image plane data in conjunction with pupil plane measurements [14], but these sets of data were not fused using Bayesian methods to estimate an image.

Much of the information from previous chapters will be repeated for clarity.

6.2   Geometry and Assumptions

The coordinate system for all the following derivations will use (x, y) functions for positions in the object plane, (α, β) in the pupil plane, (u, v) in the image plane, and (ξ, η) for shifts in the autocorrelations. However to simplify notation, we define the following variables in $\mathbb{R}^2$ to represent the ordered pairs

$$X = (x, y)$$

$$\Lambda = (\alpha, \beta)$$

$$\Psi = (\xi, \eta)$$

$$U = (u, v)$$

It also must be pointed out that for the entire paper any matrix product is a Hadamard (or element wise) product.

6.3 Image and Pupil Plane Data Models

This section is devoted to describing the statistical models for both data sets with the goal of providing probability density functions (pdf) for the measurements in both planes.

6.3.1 Image Plane Data Model. The model for the image plane intensity takes advantage of the fact that the multi-frame average of laser speckle images converges to the incoherent model [24]. This is due to the assumption that the phase at the target is random and independent from observation to observation in a manner consistent with the time varying phase distribution produced by incoherent light. By capitalizing on this, the image plane model becomes a convolution of the geometric image of the source intensity and a Point Spread Function (PSF). The PSF would include the effects of the optical system and the path turbulence. This convolution is written discretely to facilitate the derivation of a computer algorithm for image recovery.

$$i(U) = \sum_X o(X) s_i(U - X) \tag{6.1}$$

where $o(X)$ is the geometric image of the source intensity and $s_i(U)$ is the image plane PSF. Since the imaging aperture is large with respect to the Fried parameter of the turbulence, the PSF is dominated by atmospheric effects. The long exposure PSF is used since the Signal to Noise Ratio (SNR) of fully developed speckled frames is always equal to 1; making tip-tilt removal problematic.

The convolution model provides the mean of the image plane data. Since we are approximating an incoherent image using a multi-frame average, the Probability Mass Function (PMF) of the image plane data is best modeled by the PMF of partially coherent light corrupted by photon noise [19], the negative binomial distribution best describes the probability of receiving $d_i(U)$ photons shown below as

$$p_{D_i(U)}(d_i(U)|o, \mathcal{M}, s_i) = \frac{\Gamma[d_i(U) + \mathcal{M}]}{\Gamma[d_i(U) + 1]\Gamma[\mathcal{M}]} \left[1 + \frac{\mathcal{M}}{i(U)}\right]^{-d_i(U)} \left[1 + \frac{i(U)}{\mathcal{M}}\right]^{-\mathcal{M}} \tag{6.2}$$

where Γ is the gamma function, $d_i(U)$ is the measured data, $D_i(U)$ is a random variable representing the image plane data, and $\mathcal{M}$ is the number of coherent data frames used in the formation of the data, $d_i(U)$. Because the noise in each pixel is assumed to be independent and identically distributed (iid) we can write the joint pdf of the entire scene as

$$p_{D_i}(d_i) = \prod_U \frac{\Gamma[d_i(U) + \mathcal{M}]}{\Gamma[d_i(U) + 1]\Gamma[\mathcal{M}]} \left[1 + \frac{\mathcal{M}}{i(U)}\right]^{-d_i(U)} \left[1 + \frac{i(U)}{\mathcal{M}}\right]^{-\mathcal{M}} \qquad (6.3)$$

6.3.2 Pupil Plane Data Model.   The pupil plane model is based on work from Fienup and Idell [21], which relates the autocorrelation of the object brightness function to the average magnitude squared of the Fourier transform of the pupil plane intensity. The autocorrelation of the object brightness function is given by

$$R_0(\Psi) = \sum_X o(X)o(\Psi + X) \qquad (6.4)$$

Image recovery begins with estimating the average energy spectrum of the observed speckle pattern by averaging the squared moduli of many independent speckled autocorrelations

$$R_0^{(k)}(\Psi) = \mathcal{F}^{-1}\left[|\mathcal{F}[f^{(k)}(X)]|^2 H(\Lambda)\right] \qquad (6.5)$$

where $\mathcal{F}$ is a Fourier transform, f^k is the kth realization of the field reflected from the object and the aperture function, and H defines the region over which the speckle pattern is observed in the pupil plane. It has been shown that as the number of independent speckle realizations, K, increases the average energy spectrum converges to [21]

$$\lim_{K \to \infty} K^{-1} \sum_{k=1}^{K} |R_0^{(k)}(\Psi)|^2 = bs_p(\Psi) + c[R_0 * s_p](\Psi) \qquad (6.6)$$

where b and c are constants, $*$ is the convolution operator, and $s_p(\Psi)$ is the PSF of the pupil plane aperture.

The intensity in the pupil plane is related to the object field by

$$pupil^{(k)}(\Lambda) = |\mathcal{F}_{X,\Lambda}^{\lambda z}\{f^k(X)\}|^2 + n_{shot}^{(k)}(\Lambda) + n_{speckle}^{(k)}(\Lambda) \qquad (6.7)$$

where $n_{shot}^{(k)}$ is shot noise from random photon arrival times and $n_{speckle}^{(k)}$ is the speckle noise associated with coherent systems. It is convenient at this point to define a variable to represent a single frame of transformed pupil plane data

$$d_r^{(k)}(\Psi) = \left|\mathcal{F}_{\Lambda,\Psi}^{-1}\left\{|\mathcal{F}_{X,\Lambda}^{\lambda z}\{f^{(k)}(X)\}|^2 + n_{shot}^{(k)}(\Lambda) + n_{speckle}^{(k)}(\Lambda)\right\}\right|^2 \qquad (6.8)$$

where k is the frame number. According to Equation 6.5, we can average many frames of this transformed data to represent the average speckled autocorrelation of the object brightness function.

$$d_r(\Psi) = K^{-1}\sum_{k=1}^{K} d_r^{(k)}(\Psi) \qquad (6.9)$$

Provided that K is large d_r will be approximately Gaussian with mean

$$R(\Psi) = bs_p(\Psi) + c[R_0 * s_p](\Psi) \qquad (6.10)$$

This results from the central limit theorem since we are adding many independent, identically distributed random variables [6]. From this we can write an equation for the probability distribution of a single pixel

$$p_{D_r(\Psi)}(d_r(\Psi)) = \frac{exp\left\{\frac{-[d_r(\Psi)-R(\Psi)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi}\sigma} \qquad (6.11)$$

If we also assume that the noise in each pixel is independent of the noise in all the other pixels we can write the equation for the probability of realizing an entire "scene"

$$p_{D_r}(d_r|o,h) = \prod_{\Psi} \frac{exp\left\{\frac{-[d_r(\Psi)-R(\Psi)]^2}{2\sigma^2}\right\}}{\sqrt{2\pi}\sigma} \qquad (6.12)$$

6.4 Joint Algorithm

The algorithm development begins by defining a joint log-likelihood function

$$\mathcal{L}(o, b, c) = \mathcal{L}_p(o, b, c) + \mathcal{L}_i(o) \tag{6.13}$$

where

$$\mathcal{L}_p(o, b, c) = \ln p_{D_r}(d_r|o, s_p) = \sum_\Psi \frac{-[d_r(\Psi) - R(\Psi, b, c)]^2}{2\sigma^2(\Psi)} \tag{6.14}$$

where

$$R(\Psi, b, c) = bs_p(\Psi) + c[R_0 * s_p](\Psi) \tag{6.15}$$

and

$$\mathcal{L}_i(o) = \ln p_{D_i}(d_i|o, s_i, \mathcal{M}) = \sum_U d_i(U) \ln[i(U)] - [d_i(U) + M] \ln[i(U) + M] \tag{6.16}$$

The maximum likelihood (ML) estimator is defined by [34]

$$\nabla \mathcal{L}(A)|_{A=\hat{a}} = 0 \tag{6.17}$$

where $A = [o \ b \ c]^T$ and $\hat{a} = [\hat{o} \ \hat{b} \ \hat{c}]^T$

An estimate of o is difficult to find in closed form, but rather can be solved iteratively using initial estimates of b and c.

6.4.1 Estimates of b and c. To solve for initial estimates of b and c we restate the problem as the likelihood function of b and c conditioned on o

$$\mathcal{L}(b, c|o) = \sum_\Psi \frac{-[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)]^2}{2\sigma^2(\Psi)} \tag{6.18}$$

The new estimation routine is defined by

$$\nabla \mathcal{L}(A)|_{A=\hat{a}} = 0 \tag{6.19}$$

where $A = [b\ c]^T$ and $\hat{a} = [\hat{b}\ \hat{c}]^T$. These estimates will be updated as the value of o is refined. To find values of $\hat{b}$ and $\hat{c}$ we must find the gradient of $\mathcal{L}(A)$

$$\nabla \mathcal{L}(b, c) = \left[\frac{\partial \mathcal{L}(b, c)}{\partial b} \quad \frac{\partial \mathcal{L}(b, c)}{\partial c}\right]^T \tag{6.20}$$

Solving these two derivatives separately and setting them equal to zero gives us a set of simultaneous equations.

$$\frac{\partial \mathcal{L}(b, c)}{\partial b} = \sum_{\Psi} \frac{[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)]s_p(\Psi)}{\sigma^2(\Psi)} = 0 \tag{6.21}$$

$$\frac{\partial \mathcal{L}(b, c)}{\partial c} = \sum_{\Psi} \frac{[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)][R_0 * s_p](\Psi)}{\sigma^2(\Psi)} = 0 \tag{6.22}$$

Next we solve Equation 6.22 for c as a function of b

$$c = \frac{\sum_{\Psi}[d_r(\Psi) - bs_p(\Psi)][R_0 * s_p](\Psi)}{\sum_{\Psi}[R_0 * s_p]^2(\Psi)} \tag{6.23}$$

Substitute Equation 6.23 into Equation 6.21 and solve for b

$$\hat{b} = \frac{\sum_{\Psi} d_r(\Psi)s_p(\Psi) - \rho_1 \sum_{\Psi} s_p(\Psi)[R_0 * s_p](\Psi)}{\sum_{\Psi} s_p^2(\Psi) - \rho_2 \sum_{\Psi} s_p(\Psi)[R_0 * s_p](\Psi)} \tag{6.24}$$

where

$$\rho_1 = \frac{\sum_{\Psi'} d_r(\Psi')[R_0 * s_p](\Psi')}{\sum_{\Psi'}[R_0 * s_p]^2(\Psi')} \tag{6.25}$$

and

$$\rho_2 = \frac{\sum_{\Psi'} s_p(\Psi')[R_0 * s_p](\Psi')}{\sum_{\Psi'}[R_0 * s_p]^2(\Psi')} \tag{6.26}$$

Finally, substitute the value of $\hat{b}$ into Equation 6.23 to calculate a value for $\hat{c}$.

6.4.2 Estimation algorithm for o. Next we will build an iterative algorithm to solve for o. It is informative here to write the log-likelihood function to show the

65

dependance on o. We do this by recognizing

$$[R_0 * s_p](\Psi) = \sum_U \sum_X o(X)o(U+X)s_p(\Psi-U) \tag{6.27}$$

Combining this result with Equation 6.15 and (6.14) we are able to write

$$\mathcal{L}_p(o|b,c) = \sum_\Psi \frac{-[d_r(\Psi) - bs_p(\Psi) - c\sum_U \sum_X o(X)o(U+X)s_p(\Psi-U)(\Psi)]^2}{2\sigma^2(\Psi)} \tag{6.28}$$

We further must recognize

$$i(U) = \sum_X o(X)s_i(U-X) \tag{6.29}$$

Combining this result with Equation 6.16 we write

$$\mathcal{L}_i(o) = \sum_U d_i(U)\ln[\sum_X o(X)s_i(U-X)] - [d_i(U)+M]\ln[\sum_X o(X)s_i(U-X)+M] \tag{6.30}$$

This gives a joint log-likelihood of

$$
\begin{aligned}
\mathcal{L}(o|b,c) = \quad & \sum_\Psi \frac{-[d_r(\Psi) - bs_p(\Psi) - c\sum_U \sum_X o(X)o(U+X)s_p(\Psi-U)(\Psi)]^2}{2\sigma^2(\Psi)} \\
& + \sum_U d_i(U)\ln[\sum_X o(X)s_i(U-X)] \\
& - [d_i(U)+M]\ln[\sum_X o(X)s_i(U-X)+M] \tag{6.31}
\end{aligned}
$$

Next we take the derivative of this log-likelihood with respect to o

$$
\begin{aligned}
\frac{\partial \mathcal{L}(o|b,c)}{\partial o} = \quad & \sum_\Psi \frac{[d_r(\Psi) - bs_p(\Psi) - c[R_0 * s_p](\Psi)]c\sum_U \frac{\partial R_0}{\partial o}s_p(\Psi-U)}{\sigma^2(\Psi)} \\
& + \sum_U \frac{d_i(U)}{i(U)}s_i(U-X) - \frac{d_i(U)+M}{i(U)+M}s_i(U-X) \tag{6.32}
\end{aligned}
$$

where

$$\frac{\partial R}{\partial o} = o(X + U) + o(X - U) \tag{6.33}$$

Below are a few definitions to simplify this equation

$$c \sum_U \frac{\partial R_0}{\partial o} s_p(\Psi - U) = c\{[o * s_p](\Psi + X) + [o \star s_p](\Psi - X)\} \tag{6.34}$$

$$
\begin{aligned}
\Phi_1(\Psi + X) &= [o * s_p](\Psi + X) \\
\Phi_2(\Psi - X) &= [o \star s_p](\Psi - X)
\end{aligned} \tag{6.35}
$$

$$\rho(\Psi) = \frac{b s_p(\Psi) + c[R_0 * s_p](\Psi)}{\sigma^2(\Psi)} \tag{6.36}$$

Using the above definitions (and some further simplification) we are able to write a ML algorithm

$$o_{new}(X) = o_{old}(X) \frac{\left[\frac{d_i}{i_{old}} \star s_i\right](X) + c\left[\frac{d_r}{\sigma^2} \star \Phi_1\right](X) + c\left[\frac{d_r}{\sigma^2} * \Phi_2\right](X)}{\left[\frac{d_i + M}{i + M} \star s_i\right](X) + c\left[\rho \star \Phi_1\right](X) + c\left[\rho * \Phi_2\right](X)} \tag{6.37}$$

6.4.3   Stopping the Algorithm.   Image recovery techniques built in a manner similar to the Richardson-Lucy [30] algorithm are often run for a set number of iterations and then exited. It would be better to have an optimized method of exiting the iterations. It is fairly well known that iterating beyond an optimal point can lead to noise amplification [36]. This is due to the fact that all Maximum likelihood techniques attempt to fit the data as closely as possible given the constraints of the problem. This leads to the question of when do we stop our algorithm. The approach presented here involves a method designed to damp the iteration, thus avoiding noise amplification regardless of how long the algorithm is run. This means one can iterate for as long as time allows with some assurance that iteration $n + 1$ will not be worse than iteration n.

The damping routine looks at the statistics of each data set and compares it to the models statistics. When the variance of the model approaches the variance of the data, that pixel is damped. Stated mathematically for the image plane and pupil plane respectively

$$K^{-1} \sum_{k=1}^{K} (d_i^k - \hat{i})^2 < \alpha * \sigma_i^2 \tag{6.38}$$

$$K^{-1} \sum_{k=1}^{K} (d_r^k - \hat{R})^2 < \beta * \sigma_r^2 \tag{6.39}$$

where σ_i^2 and σ_r^2 are the measured variance of each data set and α and β are user chosen values to determine the degree of damping; $\hat{i}$ and $\hat{R}$ are defined as

$$\hat{i}(U) = \sum_X o_{new}(X) s_i(U - X) \tag{6.40}$$

$$\hat{R}(\Psi) = \sum_X o_{new}(X) o_{new}(\Psi + X) \tag{6.41}$$

Using the above criteria we can create two binary maps, m_i and m_r, that will damp the iteration for the pixels where the criteria are satisfied. The values of the binary maps will be set 1 for every pixel where the criteria are satisfied, otherwise the value is zero. The maps are updated at each iteration. The maps are applied to the update equation so that where ever the map equals 1 the data and the model are forced to be equal, essentially stopping that pixel from changing for that iteration. Using these criteria we restate Equations 6.36 and 6.37

$$\rho(\Psi) = \frac{bs_p(\Psi) + c \left[(1 - m_r)[R_0 * s_p](\Psi) + m_r \left[\frac{d_r - bs_p}{c} \right] \right]}{\sigma^2(\Psi)} \tag{6.42}$$

$$o_{new}(X) = o_{old}(X) \frac{\left[\left[(1 - m_i) \frac{d_i}{i_{old}} + m_i \right] \star s_i \right] (X) + c \left[\frac{d_r}{\sigma^2} \star \Phi_1 \right] (X) + c \left[\frac{d_r}{\sigma^2} * \Phi_2 \right] (X)}{\left[\left[(1 - m_i) \frac{d_i + M}{i_{old} + M} + m_i \right] \star s_i \right] (X) + c \left[\rho \star \Phi_1 \right] (X) + c \left[\rho * \Phi_2 \right] (X)} \tag{6.43}$$

where m_i and m_r are the binary maps applied to the image and pupil data sets respectively.

6.5 Results

The simulated data sets were created using Matlab® . Each set consists of 200 independently created frames of both pupil plane and image plane data. In order to create the frames of data a field magnitude was defined in the object plane. A random phase (uniform on $(-n\pi : n\pi]$) was added to every point in the field. Each phase is independent of every other phase in the field. The target was sampled at the Nyquist rate required by the pupil plane aperture. The sample rate, Δ, is calculated according to

$$\Delta = \frac{\lambda z}{2D_{pupil}} \tag{6.44}$$

where λ is the optical wavelength, z is the propagation distance, and D_{pupil} is the diameter of the aperture. This field is then propagated to the pupil plane using a Fourier transform. It can be shown that due to the random phase in the target field any propagation, regardless of propagation length, can be modeled as a Fraunhoffer propagation and therefore a simple Fourier transform. The pupil plane aperture function was then imposed on the field; after taking the magnitude squared of the field and adding photon and read noise we arrive at our raw pupil plane data. The imaging aperture function and the turbulence phase screen was then applyied to the pupil plane field and another Fourier transform was performed to give us the image plane field; the magnitude squared gives us the image plane intensity. Photon and read noise were added in the image plane.

After the data sets were created, three image recovery algorithms were run: 1) deconvolution [25], 2)Correlography [21] and 3)the joint algorithm. For each algorithm the raw image data is used as the initial estimate. It is often difficult to chose metrics to quantify the quality of a recovered image, since for almost any metric an object estimate can be chosen that is much better or worse than the chosen metric implies.

Initially mean absolute error (MAE) was chosen as a metric. MAE is defined as

$$MAE = (NM)^{-1} \sum_{n=1,m=1}^{N,M} |o - \hat{o}| \qquad (6.45)$$

This metric was quickly discarded as the errors were excessively large for estimates that were shifted with respect to the truth object. To avoid this problem we choose a slightly more complicated metric that is translation invariant [9]

$$Er = \min_{u_0,v_0} \frac{\sum |\hat{o}(u - u_0, v - v_0) - o(u, v)|^2}{\sum |o(u, v)|^2} \qquad (6.46)$$

which is more easily calculated as

$$Er = \frac{r_{\hat{o}\hat{o}}(0, 0) + r_{oo}(0, 0) - 2\max_{u_0,v_0} Re\{r_{o\hat{o}}(u_0, v_0)\}}{r_{oo}(0, 0)} \qquad (6.47)$$

where r_{oo} and $r_{\hat{o}\hat{o}}$ are autocorrelations of the object and the estimate, and $r_{o\hat{o}}$ is the cross correlation. An estimator will seek to minimize this metric, Er.

The truth target, seen in Figure 6.1 was coherently illuminated and the reflected energy was propagated along a turbulent path to a imaging system seen in Figure 6.2. The measured intensity was recorded and sent to the image recovery algorithms.

6.5.1   *Baseline Results.*   Figure 6.3 show the results of the joint algorithm compared to the results of the deconvolution algorithm for 3500 iterations with $\frac{D}{r_0} = 10$. As predicted the joint algorithm give better results at every iteration. For this comparison each algorithm was run with the same data set and using the same starting image. Figure 6.4 show the recovered image from each algorithm; clearly the joint algorithm recovers much more of the high frequency detail.

To fully quantify how well the joint algorithm works we must look at more than just a single data set. One hundred different data sets were created and the joint algorithm run for 3500 iterations on each; the results are shown in Figure 6.5. This graph demonstrates that the algorithm results have a reasonably low variance.

70
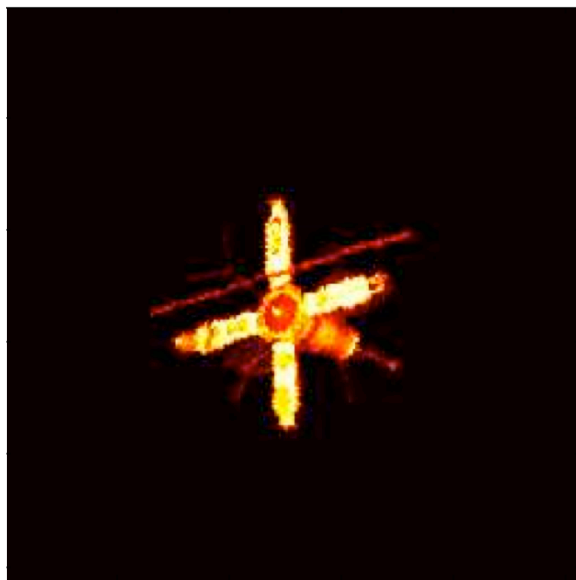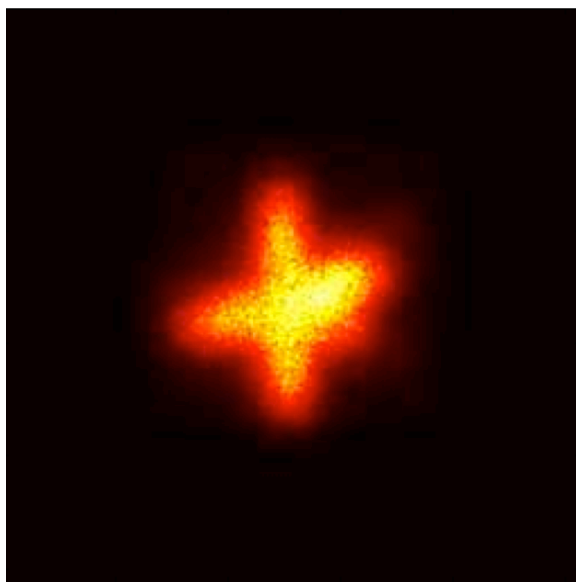
(a) Truth

(b) Raw Data

Figure 6.1:   Truth image used for all simulations and one realization of raw image data
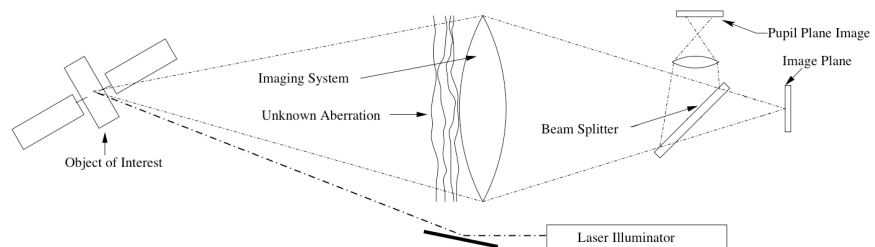
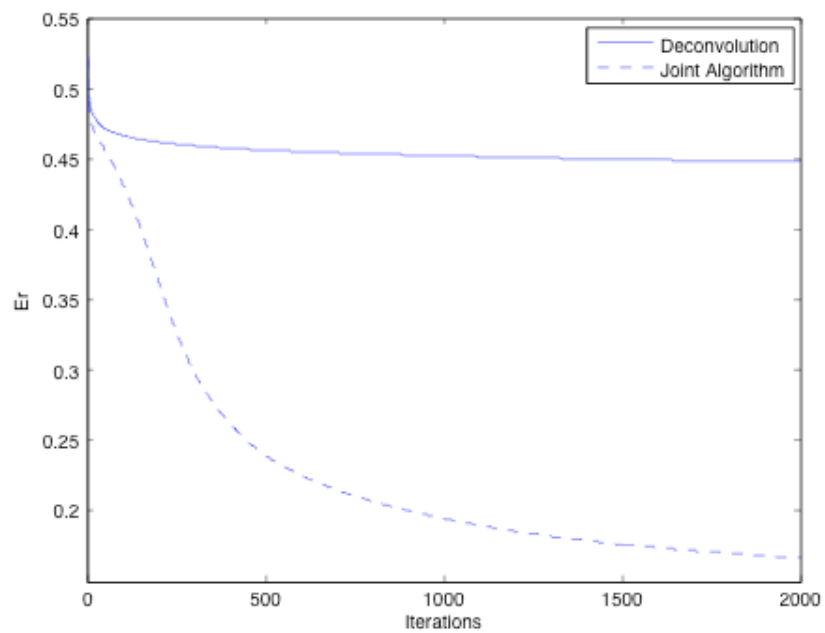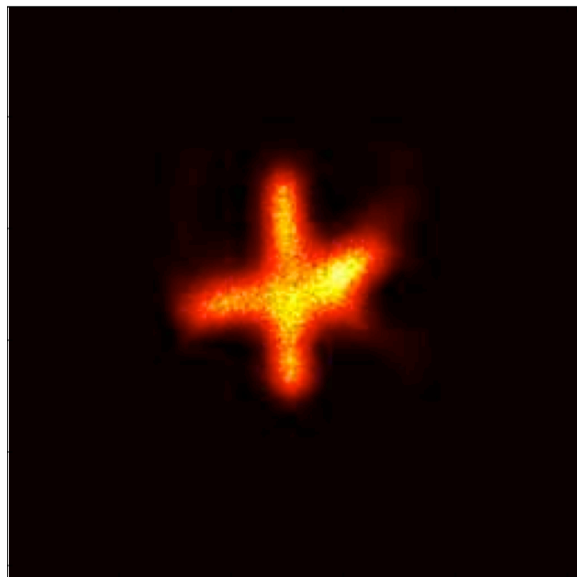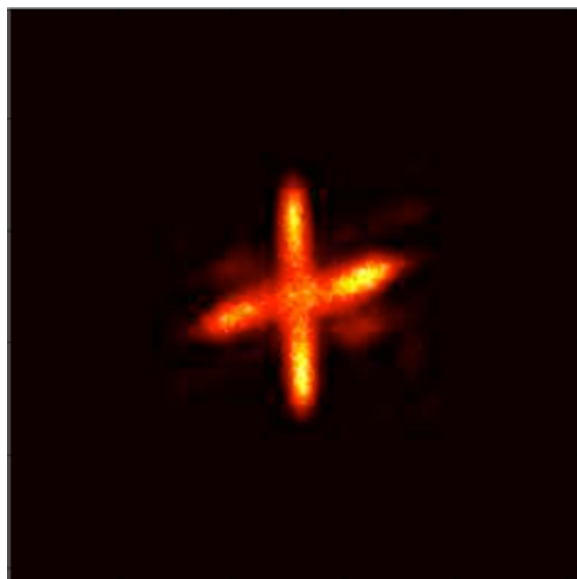Figure 6.2:   Notional system architecture

Figure 6.3:   Plot showing how the results of the joint algorithm compares to the results from deconvolution

(a) Deconvolution

(b) Joint

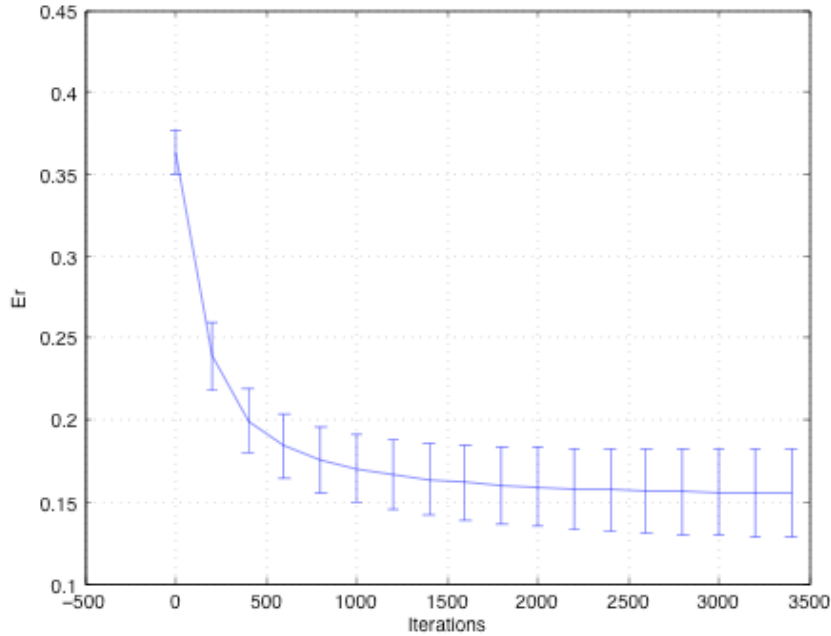Figure 6.4:   Recovered image from each algorithm at 3500 iterations

Figure 6.5: Results joint algorithm run on 100 data sets. The mean was plotted with error bars showing one standard deviation of the data.

The next step in algorithm validation was to verify that the damping of the algorithm was working properly. Figure 6.6 shows Er for 3500 iterations of both the damped and the undamped joint algorithm. It should be clear that without damping a very accurate stopping criteria would be needed. The damped algorithm removed the need for a stopping criteria at the expense of speed. Both algorithms ultimately give about the same quality of recovered image. Figure 6.7 show the recovered images at varying number of iterations; it is important to note that the image is very stable and very little change occurs between 3500 and 10000 iterations. From an application stand point the damped algorithm is run for as long as time allows without the fear of over iterating. In real time applications the undamped algorithm can be used for speed, but with the risk of iterating beyond an optimal solution.

In order to compare the results of the new algorithm to existing pupil plane method a 500 frame data set was simulated and processed using traditional imaging correlography techniques [21]. The results of these simulations are shown in Figure 6.8.
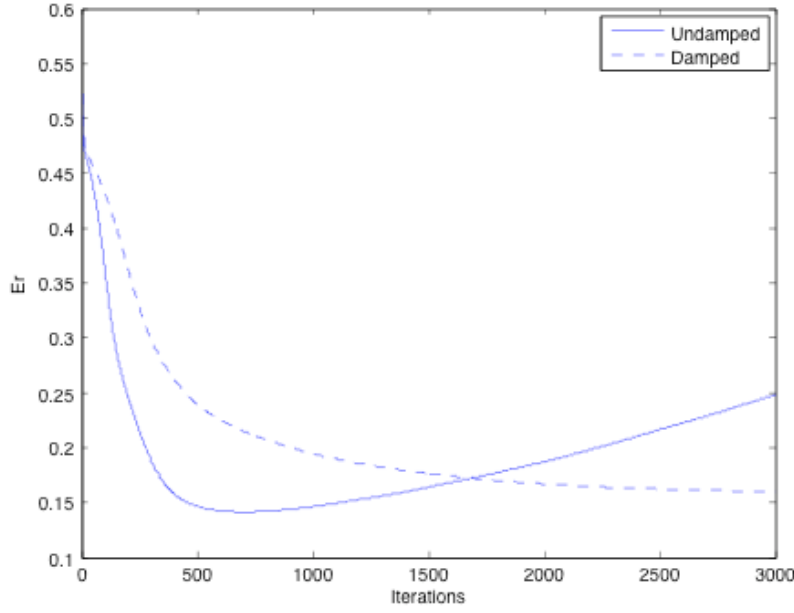
74

Figure 6.6:   Plot showing how the results of the undamped joint algorithm compare
to the results of the damped joint algorithm

Clearly the images recovered by the new algorithm have less speckle. Correlography
requires many more frames of data to overcome this speckled appearance. It should
also be noted that the new algorithm does not require a human in the loop; correl-
ography techniques require the removal of a dc term (from the PSF of the aperture)
to be removed from the data prior to image recovery, while the new algorithm does
this automatically. Further the new algorithm also provides a repeatable result by
damping the iterations.

6.5.2 Varying the number of data frames.   In order to fully understand
algorithm performance it is necessary to show results as a function of the number of
frames of data. Figure 6.9a shows Er as a function of the number of frames. There
should be two things you notice from this plot: (1) Er increases as the number of
frames decreases and (2) the damping criteria appears to break down for low frame
counts. The fact that Er increases with lower frame counts should come as no surprise;
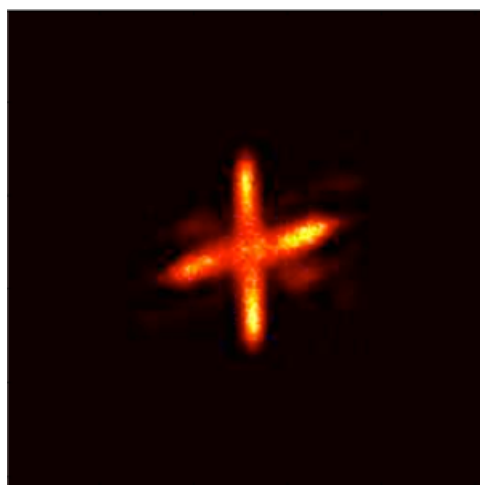our model is based on the limit as the number of frames increases without bound,
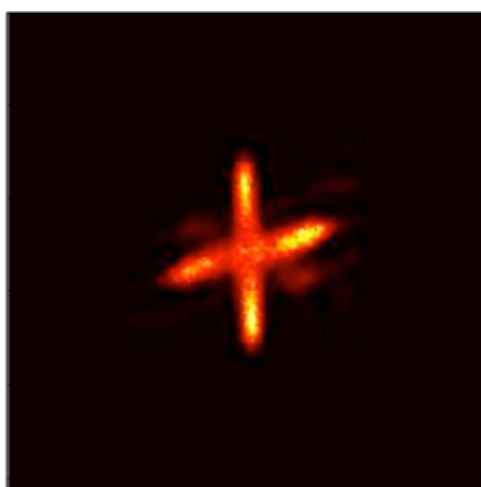
75
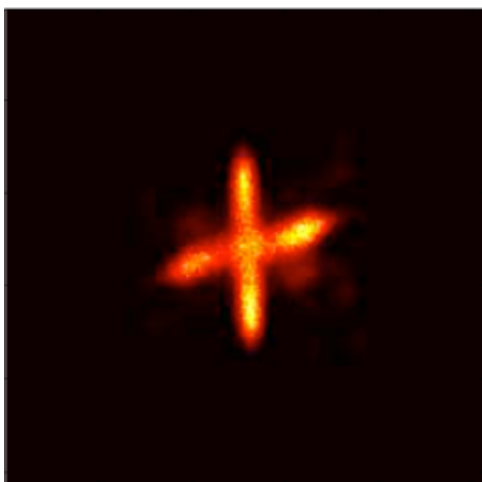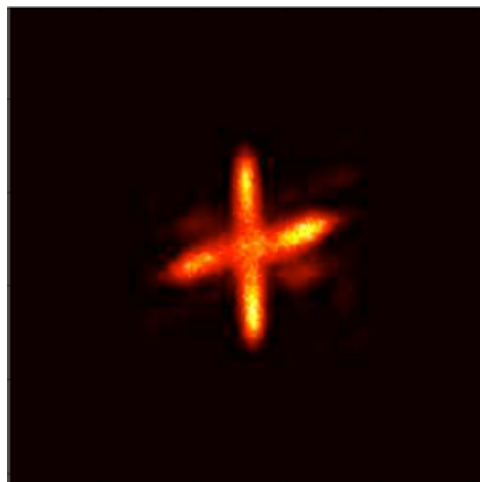
(a) 1500 Iterations

(b) 3500 Iterations
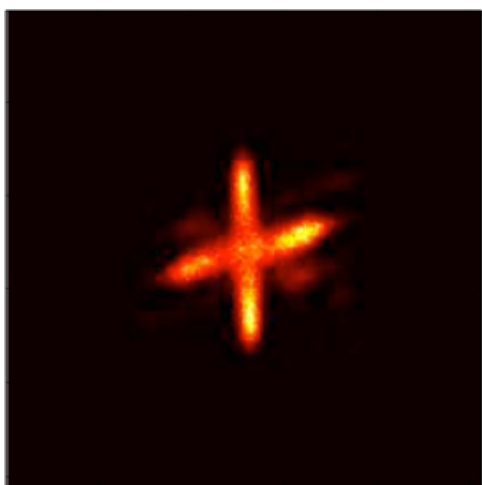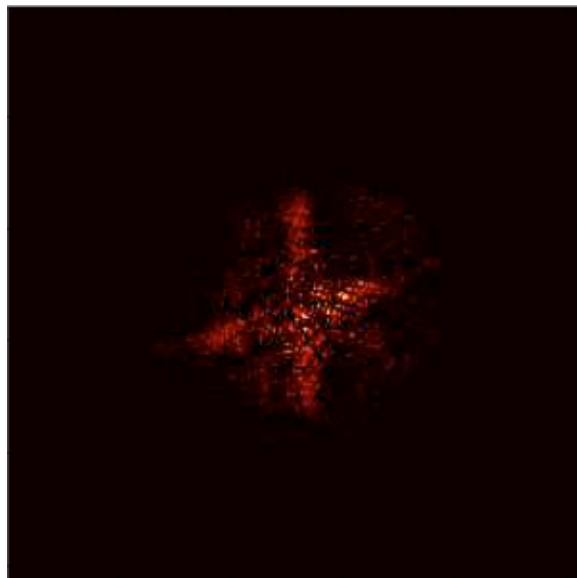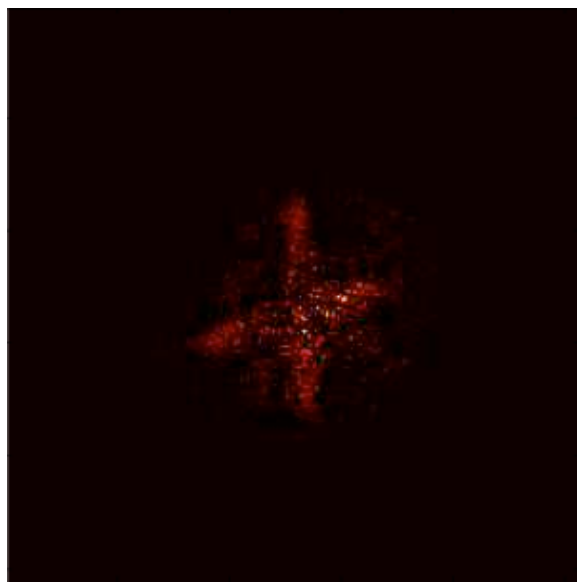

(c) 5000 Iterations

(d) 7500 Iterations

(e) 10000 Iterations

Figure 6.7:　Recovered image for varying number of iterations

(a) 1500 iterations

(b) 10000 iterations

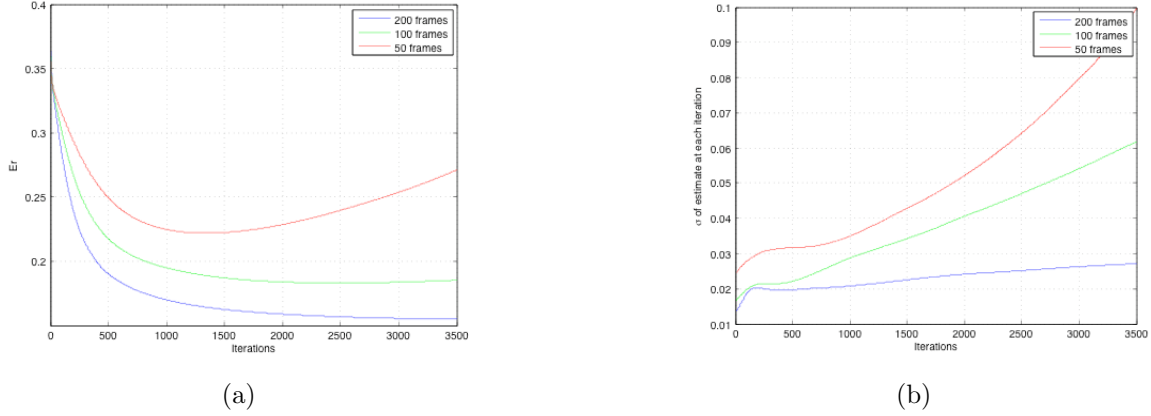Figure 6.8:   Results of imaging correlography using a 500 frame data set

Figure 6.9:   Plots showing the effects of varying the number of available data frames. (a) shows Er as a function of frames and iterations while (b) show the variance of the estimate at each iteration for different numbers of data frames

therefore more frames gives data that more closely matches the model. Also more frames will yield a higher SNR. As for the damping criteria breaking down, the method does not break down, but rather we have just not chosen the optimal value for the damping. It was hypothesized earlier in the chapter that the choice of the damping parameter would be related to the number of frames. This reinforces that thought, but demonstrates that it is not the linear relationship used. Figure 6.9b shows the variance of Er as a function of frames. Clearly the variance will rise with fewer frames of data. The reason for this is that the SNR goes down as the number of frames goes down.

6.5.3   Effects of varying the strength of the turbulence.   Since the algorithm uses both pupil and image plane data, turbulence should have some effects on the output, but not as much as a it would for a image plane only algorithm. The joint algorithm was run for four different turbulence values; the results are shown in Figure 6.10. Stronger turbulence clearly gives a higher error metric, however the difference is much less pronounced than for deconvolution. To further evaluate the effects of turbulence we turn to a subjective look at reconstructed images. Looking at Figure 6.11 we can see that the effects of turbulence are present, but the image degrades slowly with increasing turbulence.
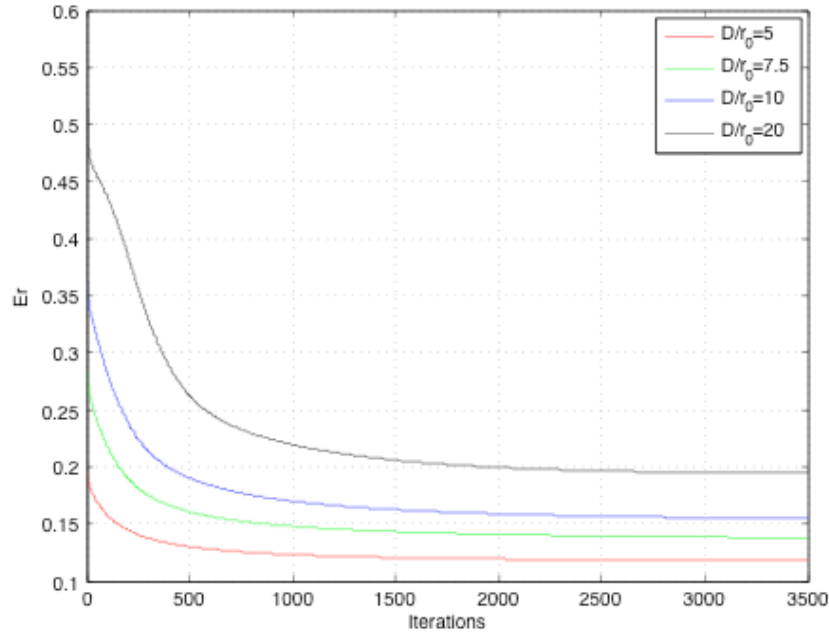
Figure 6.10: Plot showing how the results of the joint algorithm are impacted by turbulence strength

6.5.4 Ability to recover varied intensities. Figure 6.12 shows the results of running the recovery algorithm on a gray scale image. In the truth image the center area has a intensity of 1 while the other regions (moving clockwise from the upper left) have intensities of 0.1, 0.25, 0.5, and 0.75. The algorithm is not designed to recover absolute radiometry, but it is important to be able to recover relative intensity values. The outer regions in the recovered image have intensities of 10%, 22.6%, 44.1% and 70.8% given as a ratio to the center region intensity. Further the image contains greater than 90% of the energy in the field of view.

6.5.5 Quantifying system resolution. The final analysis step is to define the system resolution in a manner consistent with what was done in Chapters 4 and 5, which was to image progressively smaller bar targets and define the system resolution as the smallest separation that could be resolved. Recall that our criteria for resolution is only three bars are in the recovered image and the intensity in the valleys must be
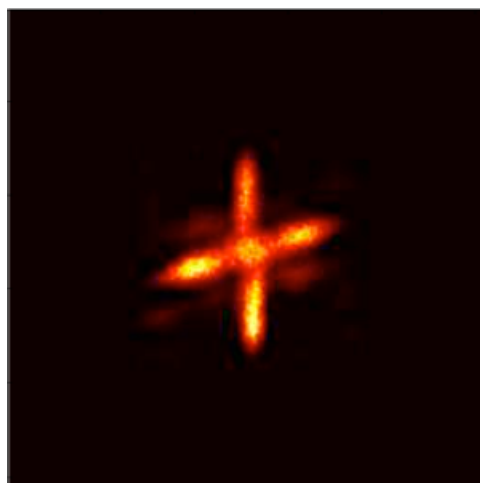
79

(a) $\frac{D}{r_0} = 5$

(b) $\frac{D}{r_0} = 7.5$
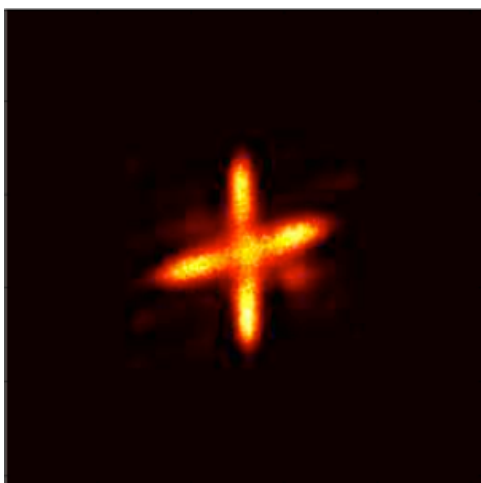
(c) $\frac{D}{r_0} = 10$

(d) $\frac{D}{r_0} = 20$

Figure 6.11:   Recovered image for varying turbulence strengths
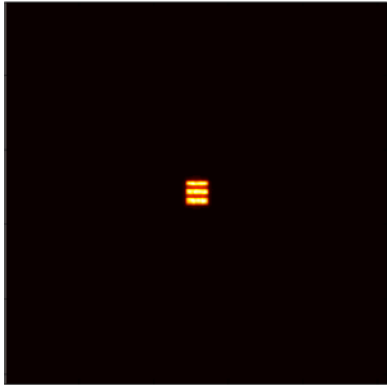
(a) Truth Object

(b) Recovered image

Figure 6.12:   Gray Scale recovered images

less than half the intensity of the peak. The results are shown in Figure 6.13; the joint algorithm was able to resolve bars separated by only three pixels.
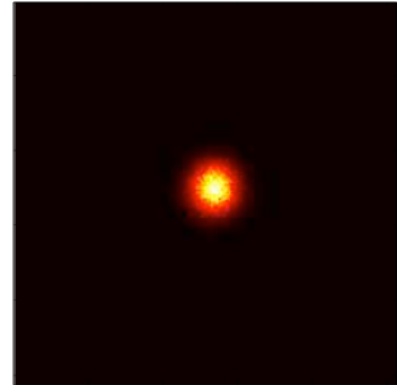
6.6   Conclusion

The joint algorithm proposed here provides an image recovery method that is less affected by turbulence than deconvolution and suffers from fewer stagnation problems than other pupil plane algorithms. The algorithm design also avoids problems such as noise amplification and image decimation commonly associated with over iterating.

This work was performed as an initial step in implementing a high resolution LADAR system using multiple data sets. In the simulations for this paper the image plane and pupil plane data sets were collected through identical apertures; this does not have to be the case. A major benefit to this algorithm is the pupil plane data can be collected using a conformal array, while the imaging portion is made small enough to be installed in current imaging platforms. This has the potential to allow for higher resolution images without greatly increasing system volume. Since the changes to the imaging chain are only software changes the pupil plane array could be added to augment existing imaging systems rather than building entirely new.
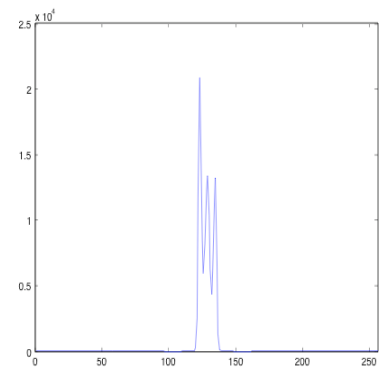
81

(a) Truth Object

(b) Raw Image Data

(c) Recovered Image

(d) Slice through the recovered image

Figure 6.13:   Quantification of the resolution of the joint algorithm

VII. Conclusion

Three image recovery algorithms for use with coherent illumination have been derived and implemented. The first was a minor modification to the stopping criteria of a previous work [23], while the other two are new work entirely. All the algorithms are based on maximum likelihood methods. Each algorithm was evaluated using Monte Carlo simulations. The improvement of the joint algorithm over the image plane (deconvolution) algorithm or pupil plane algorithm alone will provide a jump in operational capability. The design of the algorithm should also allow for easy hardware implementation.

7.1 Summary of results

In the previous chapters the results of each algorithm were analyzed and documented. The analysis for each included

1. Quantifying the resolution of each algorithm under identical conditions

2. Quantifying the ability of each algorithm to recover relative radiometry

3. Showing the effects of turbulence strength on the recovered images

4. Showing the effects of the number of data frames on the recovered images

The resolution of each algorithm was defined in terms of pixels using simple bar targets. For identical aperture sizes and turbulence parameters the resolution of each system is 10 pixels for deconvolution, 5 pixels for the pupil algorithm and 3 pixels for the joint algorithm. Figure 7.1 shows the smallest resolvable bar target for each algorithm.; clearly the gained resolution is significant.

Both deconvolution and the joint algorithm were able to recover the relative radiometry while the pupil algorithm was not. This is easily explained by the symmetry of the data model for the pupil plane data. This shows the importance of the image plane to the joint algorithm.

For deconvolution and the joint algorithm the effects of turbulence were analyzed. In both cases stronger turbulence gives a degraded reconstruction, but the
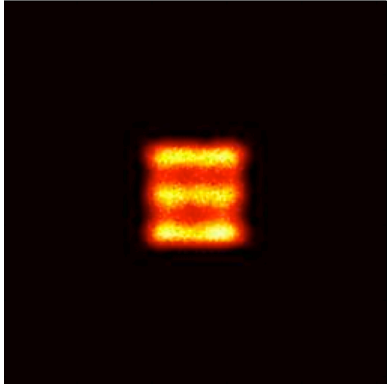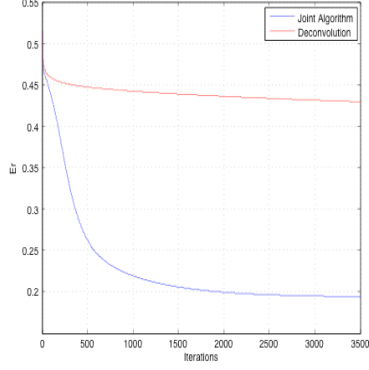
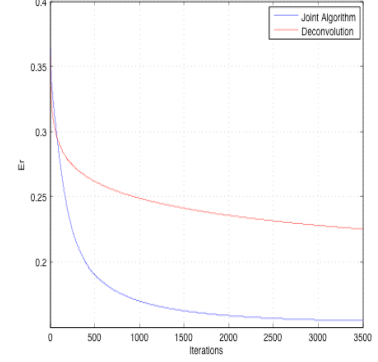(a) Deconvolution
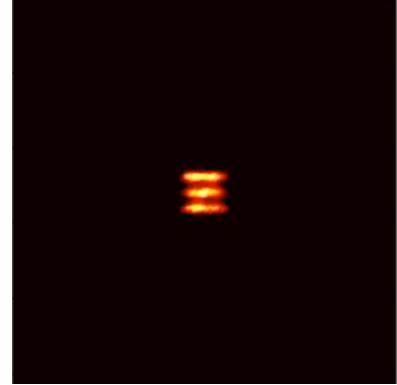
(b) Pupil Algorithm

(c) Joint Algorithm

Figure 7.1:   Comparison of resolution of each algorithm. Each sub figure shows the smallest resolvable bar target for each algorithm

degradation is much less for the joint algorithm. Figure 7.2 shows the error in the reconstruction for varying turbulence strengths. Only in the case of very week turbulence does the joint algorithm not outperform deconvolution. Although in weak

(a) $\frac{D}{r_0} = 20$

(b) $\frac{D}{r_0} = 10$
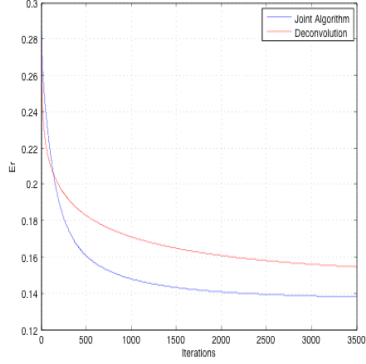
(c) $\frac{D}{r_0} = 7.5$
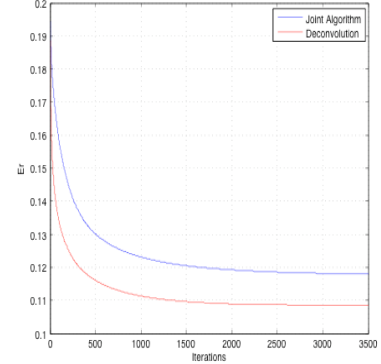
(d) $\frac{D}{r_0} = 5$

Figure 7.2:   Comparison of turbulence effects on each algorithm.

turbulence deconvolution numerically out performs the joint algorithm, a subjective look at two recovered images, Figure 7.3, shows the difference is negligible.

Finally in all cases the results are improved dramatically by adding more frames of data. This is due to the fact that both image plane and pupil plane data models are based on limit functions as the number of frames goes to infinity.

85
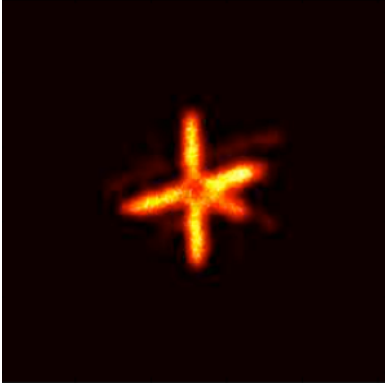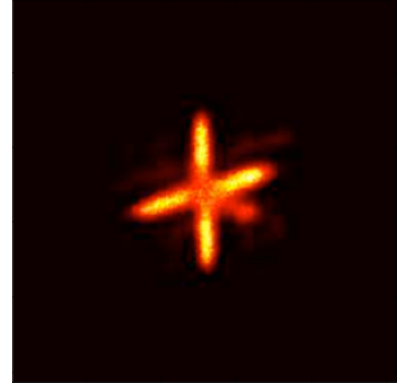
(a) Deconvolution   (b) Joint Algorithm

Figure 7.3:   Comparison of images in weak turbulence ($\frac{D}{r_0} = 5$).

7.2 Significant Contributions

1. A new more robust stopping criteria was established for use with deconvolution algorithms. The stopping criteria is based on the statistics of the data and the model and requires no *a priori* knowledge of the object being imaged.

2. A new pupil plane imaging algorithm was developed. The new algorithm has a unique stopping criteria and can deal with the dc term imposed on the data by the aperture function without a human in the loop.

3. A joint image and pupil plane algorithm was developed. This algorithm capitalizes on the strengths of both algorithms and has been shown to be more robust to turbulence than the other algorithms. The joint algorithm consistently provides better results (in terms of our chosen metric) than the other algorithms except in the case of extremely weak turbulence.

4. The joint algorithm was shown to be much less sensitive to turbulence than image plane algorithms. This was shown in both simulation and analytically (see Appendix B).

7.3 Recommended Future Work

1. Develop an optimized rule for selecting the damping parameter. This work has demonstrated some relationship between the damping parameter and the number of frames, but it is not optimized and breaks down at low frame counts.

2. Use a pupil array larger than the imaging aperture. The original intent of this effort was to form a large synthetic array; by letting the pupil array be larger than the imaging aperture it may be possible to recover additional high frequency details

3. Investigate the benefits of adding a phase diversity plane. This will likely yield little or no improvement since it will be blurred by turbulence much the same as the focal plane, however it may be interesting to look at.

4. Choose a better maximization algorithm. This algorithm seeks to maximize a likelihood function using a type of gradient accent similar to what was used by Richardson and Lucy. This may not be the best maximization algorithm; other methods may be better or faster.

7.4 Possible Applications of this Research

There are three primary application areas for this research

1. Augment existing imaging systems by adding a pupil plane detection capability

2. Build light weight imaging systems using conformal arrays

3. Implement synthetic aperture ladar systems for space based applications

Each of these application areas will be discussed briefly.

7.4.1 Augmentation of existing systems. Since the addition of the pupil plane data does not affect the image plane data, this algorithm could be applied to existing imaging systems by simply adding a pupil plane detector. This is an important consideration in terms of development cost and schedule. For example a

space observation telescope (like the Starfire Optical range) that already images space objects through the atmosphere could add a beam splitter in the optical path and re-image the pupil plane and collect the pupil data without interfering with the imaging path (other than refocusing the system). This would allow for higher resolution imagery at a lower cost than a full system redesign. One implied assumption is that the telescope is imaging laser illuminated objects. This gives an upgraded capability for Space Situational Awareness (SSA).

7.4.2 Conformal Arrays. Many of the Air Forces ISR assets are being flown on Unmanned Aerial Systems (UAS). Since these systems are small (as compared to more traditional collection platforms), there is a need for the optical systems to be light weight. This precludes the use of traditional large aperture imaging systems. Also ISR assets have to deal with turbulence so some mitigation is needed. A conformal array to collect pupil plane data, which (like the above scenario) could be used to augment the existing imager, could be added to the skin of the aircraft. This gives a larger (turbulence resistant) aperture capability will very little added weight or volume.

7.4.3 Synthetic aperture LADAR. The final application area to be discussed is synthetic aperture LADAR. This is a challenging problem due to the difficulty of measuring the phase in optical systems. The algorithm proposed by this research avoids that problem by implementing a pupil plane recovery algorithm that does not require a phase measurement. An array of sub-apertures could be formed to collect pupil intensity and this intensity used along with a low resolution image. The real benefit of synthetic aperture ladar would be in building space based systems where high resolution is needed, but it is not feasible to use large monolithic apertures.

Appendix A. Important Proofs

A.1   Introduction

This appendix is intended to provided detailed proofs of critical equations.

A.2   Equation 6.6

Prove:

$$\lim_{J \to \infty} J^{-1} \sum_{n=1}^{J} |R_n(\xi)|^2 = b|h(\xi)|^2 + cR(\xi) * |h(\xi)|^2 \tag{A.1}$$

where ξ is a two dimensional variable in the object plane.

Proof:

Given: $R_n(\xi)$ is a single frame of speckled autocorrelation data, $|h(\xi)|^2$ is the point spread function, b and c are constants and $R(\xi)$ is the true autocorrelation of the obejct brightness function.

It is helpful to notice that the left side of Equation A.1 is an ensemble average of speckled object autocorrelations. Since that phase associated with the surface roughness of the target is the only random part, and it changes each coherence time, we can rewrite this ensemble average as a time average (assuming we average over many coherence times).

$$\langle |R_n(\xi)|^2 \rangle = b|h(\xi)|^2 + cR(\xi) * |h(\xi)|^2 \tag{A.2}$$

The speckled autocorrelation, R_n can be written as

$$R_n(\xi) = [f_n(\xi, t) \star f_n(\xi, t)] * h(\xi) \tag{A.3}$$

where $\star$ and $*$ represent correlation and convolution respectively, and $f_n(\xi, t) = a(\xi)e^{j\phi(\xi,t)}$ is the object field distribution. By applying the autocorrelation theorem this can be rewritten as

$$R_n(\xi) = \mathcal{F}^{-1}\left\{ |\mathcal{F}\{f_n(\xi, t)\}|^2 H(x) \right\} \tag{A.4}$$

89

By plugging Equation A.4 in to the left side of Equation A.2 we get

$$\left\langle |\mathcal{F}^{-1}\left\{|\mathcal{F}\left\{f_n(\xi,t)\right\}|^2 H(x)\right\}|^2 \right\rangle \tag{A.5}$$

First we will expand the inner magnitude squared term

$$|\mathcal{F}\left\{f_n(\xi,t)\right\}|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(\xi_1)a(\xi_2)e^{j[\phi(\xi_1,t)-\phi(\xi_2,t)]}e^{-jkx(\xi_1-\xi_2)}H(x)d\xi_1 d\xi_2 \tag{A.6}$$

where $k = \frac{2\pi}{\lambda z}$. Next we take an inverse Fourier transform

$$\mathcal{F}^{-1}\left\{|\mathcal{F}\left\{f_n(\xi,t)\right\}|^2\right\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(\xi_1)a(\xi_2)e^{j[\phi(\xi_1,t)-\phi(\xi_2,t)]}e^{-jkx(\xi_1-\xi_2)}H(x)e^{jk\xi x}d\xi_1 d\xi_2 dx \tag{A.7}$$

Now taking the magnitude squared of this value we arrive at

$$|\mathcal{F}^{-1}\left\{|\mathcal{F}\left\{f_n(\xi,t)\right\}|^2\right\}|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(\xi_1)a(\xi_2)a(\xi_1')a(\xi_2')$$
$$e^{j[\phi(\xi_1,t)-\phi(\xi_2,t)-\phi(\xi_1',t)+\phi(\xi_2',t)]}e^{-jkx_1(\xi_1-\xi_2)}e^{jkx_2(\xi_1'-\xi_2')}$$
$$H(x_1)H^*(x_2)e^{jk\xi(x_1-x_2)}d\xi_1 d\xi_2 d\xi_1' d\xi_2' dx_1 dx_2 \tag{A.8}$$

Taking the time average yields

$$\left\langle |\mathcal{F}^{-1}\left\{|\mathcal{F}\left\{f_n(\xi,t)\right\}|^2\right\}|^2 \right\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(\xi_1)a(\xi_2)a(\xi_1')a(\xi_2')$$
$$\left\langle e^{j[\phi(\xi_1,t)-\phi(\xi_2,t)-\phi(\xi_1',t)+\phi(\xi_2',t)]} \right\rangle e^{-jkx_1(\xi_1-\xi_2)}e^{jkx_2(\xi_1'-\xi_2')}$$
$$H(x_1)H^*(x_2)e^{jk\xi(x_1-x_2)}d\xi_1 d\xi_2 d\xi_1' d\xi_2' dx_1 dx_2 \tag{A.9}$$

Looking at just the time dependent piece it we notice that

$$\left\langle e^{j[\phi(\xi_1,t)-\phi(\xi_2,t)-\phi(\xi_1',t)+\phi(\xi_2',t)]} \right\rangle = \delta(\xi_1-\xi_1',\xi_2-\xi_2') + \delta(\xi_1-\xi_2,\xi_1'-\xi_2') \tag{A.10}$$

where δ is a Dirac delta function. Taking advantage of the sifting property of the Dirac delta function, we now can write

$$\left\langle |R_n(\xi)|^2 \right\rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^2(\xi_1) a^2(\xi_1') H(x_1) H^*(x_2) e^{jk\xi(x_1 - x_2)} d\xi_1 d\xi_1' dx_1 dx_2$$
$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^2(\xi_1) a^2(\xi_2) e^{-jk(x_1 - x_2)(\xi_1 - \xi_2)}$$
$$H(x_1) H^*(x_2) e^{jk\xi(x_1 - x_2)} d\xi_1 d\xi_2 dx_1 dx_2 \quad (A.11)$$

Next we make the following substitutions

$$u = x_1 - x_2 \tag{A.12}$$

$$dx_2 = -du \tag{A.13}$$

$$w = \xi_1 - \xi_2 \tag{A.14}$$

$$d\xi_2 = -dw \tag{A.15}$$

which allows us to write

$$\left\langle |R_n(\xi)|^2 \right\rangle = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^2(\xi_1) a^2(\xi_1') d\xi_1 d\xi_1' \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x_1) H^*(x_1 - u) dx_1 e^{jk\xi u} du$$
$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^2(\xi_1) a^2(\xi_1 - w) d\xi_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x_1) H^*(x_1 - u) dx_1 e^{-jku(w - \xi)} du dw \quad (A.16)$$

Since H is the pupil function, we know (from Fourier Optics) that the autocorrelation of H is the Optical Transfer Function (OTF), $\mathcal{H}$, defined as

$$\mathcal{H}(u) = \int_{-\infty}^{\infty} H(x_1) H^*(x_1 - u) dx_1 \tag{A.17}$$

and the inverse Fourier transform of the OTF is the point spread function (PSF) times a constant

$$\mathcal{F}^{-1}\left\{\mathcal{H}(u)\right\} = K|h(\xi)|^2 \tag{A.18}$$

Using this relationship and taking advantage of the symmetry of the PSF we can write

$$\langle |R_n(\xi)|^2 \rangle = b|h(\xi)|^2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a^2(\xi_1)a^2(\xi_1 - w)d\xi_1 K|h(w - \xi)|^2 dw \qquad \text{(A.19)}$$

Finally we must recognize that

$$R(w) = \int_{-\infty}^{\infty} a^2(\xi_1)a^2(\xi_1 - w)d\xi_1 \qquad \text{(A.20)}$$

is the autocorrelation of the object intensity. Using this we can arrive at

$$\langle |R_n(\xi)|^2 \rangle = b|h(\xi)|^2 + c \int_{-\infty}^{\infty} R(w)|h(w - \xi)|^2 dw \qquad \text{(A.21)}$$

which is simply

$$\langle |R_n(\xi)|^2 \rangle = b|h(\xi)|^2 + cR(\xi) * |h(\xi)|^2 \qquad \text{(A.22)}$$

Appendix B. Algorithm insensitivity to atmosphere

B.1 Problem Statement

Given a field at a laser illuminated target, $g(X) = a(X)e^{j\phi(X,t)}$, where a is the amplitude of the field and ϕ is a random phase distributed uniformly on $(-\pi, \pi]$ that is statistically independent in both space and time. Now add a Kolmogorov phase screen at the target, $\theta(X, t)$, and aperture, $\psi(\Lambda)$, plane. Prove that the time average of the modulus squared of the Inverse fourier transform of the intensity at the receiving aperture in the Fraunhofer region is unaffected by these two phase screens.

B.2 Solution

The field in the aperture plane is found by taking a Fourier Transform of the field in the target plane:

$$\mathcal{F}_2\left\{a(X)e^{j\phi(X,t)}\right\} = \int_{-\infty}^{\infty} a(X)e^{j\phi(X,t)}e^{j\theta(X,t)}e^{-\frac{2\pi}{\lambda z}(X\Lambda)}dX \tag{B.1}$$

where θ is the phase imparted by the phase screen at the target. adding the phase screen at the aperture and taking the magnitude squared of this quantity yields

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(X_1)a(X_2)e^{j(\phi(X_1,t)+\theta(X_1,t)-\phi(X_2,t)-\theta(X_2,t))}\cdots$$
$$e^{-\frac{2\pi}{\lambda z}[(X_1-X_2)\Lambda]}dX_1 dX_2 \cdots$$
$$e^{j(\psi(\Lambda,t)-\psi(\Lambda,t))} \tag{B.2}$$

where ψ is the phase screen in the aperture. It should be clear that it has no influence in the result since

$$e^{j(\psi(\Lambda,t)-\psi(\Lambda,t))} = 1 \tag{B.3}$$

Equation B.2 represents a short exposure intensity in the aperture plane. The next step in the proof is to take the inverse Fourier transform of this quantity and then

take the modulus squared of that result

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(X_1)a(X_2)a^*(X_1')a^*(X_2')\cdots$$

$$e^{j(\phi(X_1,t)+\theta(X_1,t)-[\phi(X_2,t)+\theta(X_2,t)]-[\phi(X_1',t)+\theta(X_1',t)]+\phi(X_2',t)+\theta(X_2',t))}\cdots$$

$$e^{-\frac{2\pi}{\lambda z}[(X_1-X_2)\Lambda_1]}e^{\frac{2\pi}{\lambda z}[(X_1'-X_2')\Lambda_2]}\cdots$$

$$e^{\frac{2\pi}{\lambda z}[(\Lambda_1-\Lambda_2)\Psi]}dX_1dX_2dX_1'dX_2'd\Lambda_1d\Lambda_2 \qquad\text{(B.4)}$$

Next we take the time average of this quantity

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a(X_1)a(X_2)a^*(X_1')a^*(X_2')\cdots$$

$$\left\langle e^{j(\phi(X_1,t)+\theta(X_1,t)-[\phi(X_2,t)+\theta(X_2,t)]-[\phi(X_1',t)+\theta(X_1',t)]+\phi(X_2',t)+\theta(X_2',t))}\right\rangle\cdots$$

$$e^{-\frac{2\pi}{\lambda z}[(X_1-X_2)\Lambda_1]}e^{\frac{2\pi}{\lambda z}[(X_1'-X_2')\Lambda_2]}\cdots$$

$$e^{\frac{2\pi}{\lambda z}[(\Lambda_1-\Lambda_2)\Psi]}dX_1dX_2dX_1'dX_2'd\Lambda_1d\Lambda_2 \qquad\text{(B.5)}$$

It was possible to move the time average operation inside the integrals due to the fact that both are linear operations and only ϕ and θ have a time dependence.

To eliminate the effects of the phase screen at the target we need to show that the composite random variable, $z = \phi + \theta$, is uniform and uncorrelated. The first step is to write the joint pdf of z at two points in space. If this pdf is uniform and has a no spatial correlation we have shown that θ does not change the phase distribution.

$$p_z(z_1, z_2) = \int\int p_\theta(z_1 - \phi_1, z_2 - \phi_2)p_\phi(\phi_1, \phi_2)d\phi_1 d\phi_2 \qquad\text{(B.6)}$$

We were given that ϕ is uniform on $(-\pi, \pi]$ and we know the phase screens have joint Gaussian statistics and are defined for all values of phase. This makes it necessary to redefine the density of ϕ to cover all phase values. We can simplify a little by recognizing that the area under a Gaussian curve becomes very small in the tails; this

allows us to define the uniform density function over a finite range of phase

$$p'_\phi(\phi_1, \phi_2) = \frac{1}{2n\pi} \tag{B.7}$$

where $\phi_1 \in (-n\pi, n\pi]$ and $\phi_2 \in (-n\pi, n\pi]$ where n is an integer. The value of n is chosen such that

$$\int_{-n\pi}^{n\pi} \int_{-n\pi}^{n\pi} p_\theta(\theta_1, \theta_2) \approx 1 \tag{B.8}$$

By plugging eq(B.7) into eq(B.6) it can be shown that

$$p_z(z_1, z_2) = \frac{1}{2n\pi} \int_{-n\pi}^{n\pi} \int_{-n\pi}^{n\pi} p_\theta(z_1 - \phi_1, z_2 - \phi_2) d\phi_1 d\phi_2 \tag{B.9}$$

If we plug in $\theta_1 = z_1 - \phi_1$, $\theta_2 = z_2 - \phi_2$, $d\phi_1 = -d\theta_1$, and $d\phi_2 = -d\theta_2$ we can write

$$p_z(z_1, z_2) = \frac{1}{2n\pi} \int_{-n\pi}^{n\pi} \int_{-n\pi}^{n\pi} p_\theta(\theta_1, \theta_2) d\theta_1 d\theta_2 \tag{B.10}$$

which gives us

$$p_z(z_1, z_2) \approx \frac{1}{2n\pi} \tag{B.11}$$

which is clearly uniform and uncorrelated.

Appendix C. Investigation of the spatial independence of speckle noise

For speckle noise to be truly independent from pixel to pixel the average speckle size must be smaller than a pixel. We have to test this assumption in both the image plane and the pupil plane. We will look first at the pupil plane.

Dr. Joseph Goodman developed a measure for the average size of a speckle lobe in any non-imaged plane [18]. By assuming the object has uniform brightness one can find the average area of a speckle lobe by

$$S_c = \frac{\lambda^2 z^2}{A} \tag{C.1}$$

where λ is the wavelength of the optical radiation, z is the propagation length, and A is the area of the target. If we lift the uniform brightness restriction the expression is

$$S_c = \lambda^2 z^2 \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I^2(u,v) du dv}{\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(u,v) du dv \right]^2} \tag{C.2}$$

The above equations give an expression for the area of a speckle lobe. There is no expression for the linear dimensions of a speckle lobe, but taking the square root of the area will yield a good approximation.

The speckle size in the image plane is found by assuming the pupil aperture is a uniformly bright source and using the image distance as the propagation distance.

For the geometry used in this research, the speckle correlation extends over a few pixels (3-5), so the assumption of independence is not strictly true, but does not depart from it significantly.

Bibliography

1. Andrews, Larry C. and Ronald L. Phillips. *Laser Beam Propagation through Random Media.* SPIE Press, Bellingham, Washington, 1998.

2. Bracewell, Ronald N. *The Fourier Transform and its Applications.* McGraw Hill, Boston, MA, 2000.

3. Cain, Stephen C. "Joint Blind Deconvolution and Image Correlography via a Bayesian Image Reconstruction Algorithm". *Conference Proceedings*, 2006.

4. Crimmins, T.R. and J.R. Fienup. "Uniqueness of phase retrieval for functions with sufficiently disconnected support". *J. Opt. Soc. Am.*, 73(2):218–221, February 1983.

5. Crimmins, T.R., J.R. Fienup, and B.J. Thelen. "Improved bounds on object support from autocorrelation support and application to phase retrieval". *J. Opt. Soc. Am.*, 7(1):3–13, January 1990.

6. Dougherty, Edward R. *Random Processes for Image and Signal Processing.* SPIE Press, 1999.

7. Fienup, J.R. "Reconstruction of an object from the modulus of its Fourier transform". *Optics Letters*, 3(1):27–29, July 1978.

8. Fienup, J.R. "Phase retrieval algorithms: a comparison". *Applied Optics*, 21(15):2758–2769, August 1982.

9. Fienup, J.R. "Invariant error metrics for image reconstruction". *Applied Optics*, 36(32):8352–8357, November 1997.

10. Fienup, J.R. "Comparison of Reconstruction Algorithms for Images from Sparse-Aperture Systems". *SPIE Conference Proceedings*, 2002.

11. Fienup, J.R. "Lenless Coherent imaging by phase retrieval with and illumination pattern constraint". *Optics Express*, 14(2):498–508, January 2006.

12. Fienup, J.R., T.R. Crimmins, and W. Holsztynki. "Reconstruction of the support of an object from the support of its autocorrelation". *J. Opt. Soc. Am.*, 72(5):610–624, May 1982.

13. Fienup, J.R. and Paul S. Idell. "Imaging Correlography with Sparse Arrays of Detectors". *Opt. Engr.*, 27:778–784, 1988.

14. Fienup, J.R. and A. M. Kowalczyk. "Phase retrieval for a complex-valued object by using a low-resolution image". *J. Opt. Soc. Am.*, 7(3):450–458, March 1990.

15. Fienup, J.R. and C.C. Wackerman. "Phase-retrieval stagnation problems and solutions". *J. Opt. Soc. Am.*, 3(11):1897–1907, November 1986.

16. Gaskill, Jack D. *Linear Systems, Fourier Transforms and Optics*. Wiley, New York, 1978.

17. Gerchberg, R.W. and W.O. Saxon. "A Practical algorithm for the determination of phase from image and diffraction plane pictures". *Optik*, 35:237–246, 1972.

18. Goodman, Joseph. *Speckle Phenomena: Theory and Applications*. Roberts and Company, 2007.

19. Goodman, Joseph W. *Statistical Optics*. Wiley, New York, 1985.

20. Goodman, Joseph W. *Fourier Optics*. McGraw Hill, Boston, 1996.

21. Idell, Paul S., J.R. Fienup, and Ron S. Goodman. "Image Synthesis from nonimaged laser-speckle patterns". *Optics Letters*, 12(11):858–860, November 1987.

22. Johnson, Peter. *Phase Diversity and Polarization Augmented Techniques for Active Imaging*. Ph.D. thesis, Air Force Institute of Technology, 2007.

23. MacDonald, Adam. *Blind Deconvolution of Anisoplanatic Images Collected by a Partially Coherent Imaging System*. Ph.D. thesis, Air Force Institute of Technology, 2006.

24. MacDonald, Adam, S.C. Cain, and E.A. Armstrong. "Image Restoration Techniques for Partially Coherent 2-D LADAR". *Proc SPIE Conference, Denver Colorado*, 5562-10, August 2004.

25. MacDonald, Adam, S.C. Cain, and E.A. Armstrong. "MAP Image and Seeing Condition Estimation from Partially Coherent 2-D LADAR Images". *Optical Engineering*, 45:116001, 2006.

26. Magee, Eric. *Phase Screens for Long Time Series Wave Optics Simulations*. Technical report, Mission Research Corporation, Dayton, OH, May 2003.

27. Paxman, R.G., J.R. Fienup, and J.T. Clinthorne. "Effect of Tapered Illumination and Fourier Intensity Errors on Phase Retrieval". *SPIE Conference Proceedings*, 1987.

28. Petrou, Maria and Panagiota Bosdogianni. *Image Processing: The fundamentals*. John Wiley and Sons, New York, 2004.

29. Poor, H.Vincent. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, 1994.

30. Richardson, W. H. "Bayesian-Based Iterative Method of Image Restoration". *J. Opt. Soc. Am.*, 62:55–, 1972.

31. Roggeman, Michael C. and Byron M. Welsh. *Imageing Through Turbulence*. CRC Press, Boca Raton, Florida, 1996.

32. Schulz, Timothy J. "Image Recovery from Correlations". *J. Opt. Soc. Am.*, 9(8):1266–1272, August 1992.

33. Seldin, J.H. and J.R. Fienup. "Iterative blind deconvolution algorithm applied to phase retrieval". *J. Opt. Soc. Am.*, 7(3):428–433, March 1990.

34. VanTrees, H.L. *Detection, Estimation, and Modulation Theory Part I.* Wiley Interscience, New York, 2001.

35. Vardi, Y. and L. A. Shepp. "Maximum Likelihood Reconstruction in Positron Emission Tomography". *IEEE Trans. Medical Imaging*, 1(2):113–122, 1982.

36. White, R.L. "Image Restoration Using the Damped Richardson-Lucy Iteration". *Astronomical Data Analysis Software and Systems III, A.S.P Conference Proceedings*, 1994.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE		3. DATES COVERED *(From — To)*
March 27, 2008	Ph.D. Dissertation		Mar 2003 — Dec 2007

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
	DACA99–99–C–9999
Joint Image and Pupil Plane Reconstruction Algorithm based on Bayesian Techniques	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
James D Phillips, Maj, USAF	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Institute of Technology Graduate School of Engineering and Management 2950 Hobson Way WPAFB OH 45433-7765	AFIT/DEE/ENG/08-07

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approval for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The focus of this research was to develop an joint pupil and focal plane image recovery algorithm for use with coherent LADAR systems. The benefits of such a system would include increased resolution with little or no increase in system weight and volume as well as allowing for operation in the absence of natural light since the target of interest would be actively illuminated. Since a pupil plane collection aperture can be conformal, such a system would also potentially allow for the formation of large synthetic apertures. The system is demonstrated to be robust and in all but extreme cases yield better results than algorithms using a single data set (such as deconvolution). It was shown that the joint algorithm had a resolution increase of 70% over deconvolution alone and a 40% increase over traditional pupil plane algorithms. It is also demonstrated that the new algorithm does not suffer as severely from stagnation problems typical with pupil plane algorithms.

15. SUBJECT TERMS

synthetic aperture ladar, image recovery, deconvolution, phase retrieval, correlography

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr Stephen Cain
U	U	U	UU	112	19b. TELEPHONE NUMBER *(include area code)* (937) 255–3636 email: stephen.cain@afit.edu